A Quasi-Bayesian Local Likelihood Approach to Time Varying Parameter Models

Katerina Petrova

A thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy (Ph.D.) in Economics

School of Economics and Finance
Queen Mary University London

September 5, 2016

Declaration of Authorship

I, Katerina Petrova, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to my best knowledge break any UK law, infringe any third party's copy right or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: September 5, 2016

Details of collaborations and publications:

Chapter 3 is working paper, joint with George Kapetanios, Liudas Giraitis and Ana Galvão.

Chapter 4 is a joint paper with George Kapetanios, Liudas Giraitis and Ana Galvão, published in the *Journal of Empirical Finance.*

Chapter 5 is a working paper, joint with George Kapetanios, Riccardo Masolo and Matthew Walderon.

The algorithm in Chapter 3.3.2 is in a working paper joint with Ana Galvão.

Acknowledgements

I would like to thank George Kapetanios for giving me the opportunity to work alongside him, for always trusting me with important projects and never micro-managing me. It is precisely this that taught me to be a responsible and independent researcher.

There are no words to describe how much I owe to Liudas Giraitis, he is the kindest and most generous person I have known and his door has always been open for me.

I am indebted to Ana Galvão for everything she taught me and for being as strict as a parent. She taught me discipline and it is because of her insistence that I completed my job market paper on time for job applications. And, of course, it was her course I took in my MSc., that got me interested in macroeconometrics in the first place.

I would also like to thank Haroon Mumtaz for all the valuable comments and discussions and for being a good friend.

Finally, I would like to thank Tassos for his infinite love and patience throughout the writing of this work.

Abstract

This thesis proposes a new econometric methodology for the estimation and inference of macroeconomic models in the presence of time variation in the parameters. A novel quasi-Bayesian local likelihood (QBLL) approach is established and it is shown that the method gives rise to asymptotically valid quasi-posterior distributions. In addition, in the special case of linear Gaussian models, expressions of the quasi-posteriors are derived in closed form, which simplifies inference and makes the use of MCMC unnecessary. Inference based on the QBLL approach, as a consequence of modelling parameter variation nonparametrically, is robust to different processes for the drifting parameters, as its validity does not depend on parametric restrictions typically imposed by alternative state space models. In addition, the Bayesian treatment of the approach provides a remedy to the 'curse of dimensionality' by accommodating large dimensional systems. We demonstrate that the proposed estimators exhibit good finite sample properties, and, unlike the alternative parametric state space models, are robust to different parameter processes. We provide a variety of interesting macroeconomic applications and forecasting exercises to reduced-form VAR models. In addition, we develop the methodology to the estimation of structural DSGE models in the presence of parameter drift. We apply the proposed algorithms to different medium-sized DSGE models in order to study structural change in the parameters.

Contents

List of Tables

List of Figures

1 Introduction

Standard econometric techniques usually assume that the data generating process depends on a number of fixed parameters. In a macroeconomic context, this has the implication that relationships between economic variables remain constant over time. This assumption is not supported by evidence from macro time series, where relationships between variables are subject to structural change: in the case of the recent 2008 financial crisis this structural change was abrupt, while in the context of the transition of the world economy from the volatile period of the 1970s-1980s oil crises to the low volatility period of the Great Moderation in the late 1980s and 1990s, the structural change was slow and gradual. Standard econometric models fail to capture such time varying relationships and deliver invalid inference in the cases where time variation is erroneously ignored. The issue has assumed increased practical relevance as time series samples span longer periods due to the increase of data availability, so the problem of accommodating time varying relationships has assumed a prominent role in econometric research.

This thesis contributes to this research by combining existing nonparametric approaches with Bayesian methods and establishing a quasi-Bayesian local likelihood (QBLL) estimation methodology for a general multivariate model with time varying parameters. The proposed estimators differ from existing state space approaches in that they estimate parameter time variation nonparametrically without imposing assumptions on the stochastic processes of the parameters. The QBLL approach augments the frequentist estimators of Giraitis, Kapetanios and Yates (2014) and Giraitis, Kapetanios, Wetherilt and Zikes (2016) in order to provide a Bayesian treatment for the drifting parameters. It is the Bayesian treatment that allows to increase the dimension of the models and helps avoid overfitting. This is particularly relevant for Vector Autoregressions (VAR) where the number of parameters is large. In addition, the Bayesian treatment enables the construction of MCMC algorithms that can sample from the joint posterior of the parameters in the presence of mixtures of time varying and time invariant parameters.

In Chapter 2, we establish the QBLL methodology and prove that the resulting quasi-posterior distributions are asymptotically valid for inference and confidence interval construction. In the special case of linear Gaussian models, we derive closed form expressions for the quasi-posterior densities: these are of Normal-Gamma and Normal-Wishart distributional form, which alleviates the need to use MCMC methods and makes the approach fast and tractable. Chapter 2 also positions the method in the existing literature, comparing and contrasting it to: i) existing parametric state space approaches and ii) nonparametric approaches. Another contribution of Chapter 2 is to constructs several Gibbs sampling algorithms, which can sample from the joint posterior distribution of combinations of time varying and time invariant parameters. This is done in line with the nonparametric strategy of this thesis, hence without imposing parametric assumptions on the parameters. In addition, in Chapter 2 we provide a study of the finite sample performance of the QBLL estimators and compare them to alternative methods. Finally, we demonstrate the method in action with an empirical application and a forecasting exercise using Bayesian VAR models with time varying parameters.

Having established the theoretical framework and applied it to reduced form models in Chapter 2, we turn to structural models in Chapter 3. In particular, we develop further the approach and demonstrate how it can be employed for the estimation of dynamic stochastic general equilibrium (DSGE) models in the presence of drifting parameters. DSGE models have recently received considerable attention in policy analysis and forecasting and we demonstrate that allowing for parameter variation can not only be very effective in detecting structural change or model misspecification, but it can also significantly improve the forecasting performance of the model, which can have wide ranging policy applications.

The methodology of Chapter 3 is general and can be applied to any DSGE model and Chapter 4 and 5 are applications of the approach to different DSGE models. In Chapter 4, we address an important drawback of DSGE models after the events of 2007-8, namely the lack of a financial sector and as a result, the inability of standard DSGE models to fit well the data in-sample or forecast out-of-sample after the beginning of the financial crisis. To this end, Chapter 4 estimates a DSGE model with financial frictions with the QBLL approach. After allowing for parameter drift, we find that the parameters guiding the financial sector in the model do not change after 2008; instead, the volatility of the financial shock doubles in the crisis, suggesting a new interpretation of the recent crisis as a 'Bad Luck' event.

In Chapter 5, we apply the QBLL approach to a medium-sized open economy DSGE model for the UK. The model is known as COMPASS and is the main organising model for policy analysis and forecasting in the Bank of England. We find that the relative importance of the risk premium shock of the model, measured as the wedge between the policy instrument and the interest rate that households and firms face, has increased during the 2008 crisis, indicating changes in the credit availability.

Finally, Chapter 6 summarises the findings of the thesis and provides concluding remarks. The Appendix in Chapter 7 contains the proofs of all propositions, algorithm descriptions and some additional results.

2 The QBLL approach

2.1 Introduction

Time varying parameter models have recently received a lot of attention in macroeconometric literature, due to their ability to accommodate structural changes or breaks in the relationships between key macroeconomic variables. As the time span of available data increases, it is implausible that relationships between economic variables remain fixed; consequently, the issue of accommodating parameter time variation in econometric models, such as vector autoregressions (VARs), has assumed an increasingly prominent role both in theoretical and empirical macroeconometric research.

The standard approach to the estimation of time varying parameter (TVP) VAR models employs state space methods. In a seminal paper, Cogley and Sargent (2002) study the changing dynamics of macroeconomic variables in the U.S. using a TVP-VAR with autoregressive coefficients modelled as random walk processes. In subsequent work, Sims (2001) and Stock (2001) note that the uncovered parameter drift reported in Cogley and Sargent (2002) may be exaggerated as a result of a time invariant covariance matrix assumption. Primiceri (2005) and Cogley and Sargent (2005) address this issue by accommodating drifting volatility in VAR models, the former by employing a procedure suggested by Kim, Shephard and Chib (1998) and the latter by utilising a Metropolis within Gibbs algorithm with a technique from Jacquier, Polson and Rossi (1994). Both approaches require a parametric specification of the stochastic process generating the volatility parameters. More recently, Cogley, Primiceri and Sargent (2010) propose a VAR model, which in addition to drifts in the parameters and volatilities, also features time varying volatility in the state equations of the autoregressive parameters.

An alternative to the parametric state space approach is presented in Giraitis et al. (2014), who propose a nonparametric method for the estimation of the coefficient and variance processes in a time varying linear regression setting and establish the theoretical properties of their kernel-type estimator. Their method is extended to a general local likelihood framework that accommodates consistent and asymptotically normal extremum estimation by Giraitis et al. (2016).

Neither state space models nor the frequentist estimators of Giraitis et al. (2014, 2016) can accommodate large dimensional systems in the presence of time variation in the parameters. This Chapter attempts to address this issue by combining the existing nonparametric approach with Bayesian methods and establishing a quasi-Bayesian local likelihood (QBLL) estimation methodology for a general multivariate model with time varying parameters. The QBLL approach augments the frequentist estimators of Giraitis et al. (2016) in order to provide a Bayesian treatment for the drifting parameters and avoid over-fitting. We prove that the resulting quasi-posterior distributions are asymptotically valid for inference and confidence interval construction. In addition, we derive in closed form a Normal-Wishart expression for the quasi-posterior density in the special case of a Gaussian VAR model.

Another key issue addressed in this Chapter is the development of an econometric procedure that can estimate mixtures of time invariant and time varying parameters, by developing several Gibbs algorithms based on the analytic expressions obtained for the conditional quasi-posterior densities. It is the Bayesian treatment of this Chapter that facilitates the construction of such algorithms and to our best knowledge, this is the first procedure that can accommodate such mixtures without imposing parametric assumptions on the parameter processes.

The proposed QBLL method has several advantages over the widely used state space approaches to drifting parameters. First, the parameter time variation enters nonparametrically, ensuring consistent estimation in a wide class of parameter processes, and alleviating the risk of invalid inference due to misspecification of the state equation. This point is illustrated further in the Monte Carlo exercise. Second, unlike standard state space methods, the proposed method does not suffer from dimensionality issues. The combination of our closed form quasi-posterior expressions with Minnesota-type priors, specified directly on the drifting parameters, allows the number of variables in the VAR to be very large. To illustrate this point, the QBLL approach can handle estimation of an 87 variable TVP-VAR model (see Section 2.10.2), whereas state space methods are limited to models with at most 4 variables. The availability of analytic expressions for the joint quasi-posterior density in the Gaussian VAR case alleviates the computational burden of MCMC methods used for the estimation of TVP-VARs by state space methods. Another advantage of the proposed quasi-Bayesian approach is the inverted-Wishart property of the time varying covariance matrix, which ensures symmetric positive definiteness of the resulting estimators, without the need of further restrictions.

We apply our novel QBLL approach to study the changing dynamics in a VAR model with key U.S. macroeconomic variables. We find considerable time variation in the series for the natural rate of unemployment and core inflation, as well as a significant decline in inflation persistence after the beginning of the Great Moderation period. These results confirm evidence in Cogley and Sargent (2002, 2005) on the time varying dynamics of inflation. We also uncover a fall in the volatilities of the series after the end of the oil crises, which is in line with evidence presented in Primiceri (2005) and Sims and Zha (2006).

In order to access the forecast record of the QBLL estimators, we design an out-of-sample forecasting exercise. We find that allowing for time varying parameters can improve both point and density forecasts of BVAR models. Increasing the dimension of the VAR along with parameter time variation, which is made feasible by the use of the QBLL methodology established in this Chapter, can enhance forecast performance further, delivering unbiased forecasts and uniform probability integral transformations. Moreover, the algorithms developed in this Chapter, by accommodating mixtures of time varying and time invariant parameters, allow us to 'switch off' variation in the autoregressive component or in the volatility of the VAR and investigate further the source of forecast improvements. To this end, we find that drifts in the autoregressive matrix can eliminate forecast bias while variation in the volatility improves density forecasts.

The rest of the Chapter is organised as follows. Section 2.2 develops the QBLL methodology and establishes its asymptotic validity. In Sections 2.3 and 2.4, we derive closed form quasi-posterior expressions for the special case of a linear Gaussian regression model and a VAR model respectively; these expressions provide the building block of a number of Gibbs algorithms developed in Section 2.8. In Section 2.9, we study the finite sample properties of the QBLL estimators and compare them to existing methods in a Monte Carlo experiment. Section 2.10 contains our empirical contribution to the literature on changing dynamics of inflation and monetary policy in the US and the forecasting exercise. Section 2.11 provides a concluding discussion and Appendix 7.1 contains all proofs of propositions and some additional algorithms and empirical results.

2.2 Asymptotic Theory

This Section establishes the QBLL methodology for inference in the presence of time varying parameters. Bayesian analysis is conducted by employing a modification to the Giraitis et al. (2016) frequentist estimator, augmenting the resulting modified objective function by a prior density and deriving a posterior density for the purpose of estimation and confidence interval construction. The asymptotic validity of the posterior density arising from the QBLL approach is formally established by Proposition 1 below.

Let y_t be an observed time series with log-density $l_t(y_t|y^{t-1}, \theta_t)$, conditional on history $y^{t-1} = \{y_1, ..., y_{t-1}\}$, depending on a time varying finite-dimensional vector of parameters θ_t, satisfying one of the conditions (i)-(ii) presented below.

(i) For each $t \in \{1, ..., T\}$, θ_t is a deterministic function of time given by

$$\theta_t = \theta\left(\frac{t}{T}\right) \tag{1}$$

where $\theta(.)$ is a piecewise differentiable function.

(ii) For $1 \leq h \leq t$ as $h \to \infty$, θ_t is a vector-valued stochastic process satisfying

$$\sup_{j:|j-t|\leq h} ||\theta_t - \theta_j||^2 = O_p\left(h/t\right). \tag{2}$$

Both (1) and (2) imply that the sequence of parameters drifts slowly with time, a property that is important for consistent estimation of θ_t. An extremum estimator $\hat{\theta}_j = \arg\max_\theta \ell_j(\theta_j)$ for θ_j is derived by maximising an objective function given by

$$\ell_j(\theta_j) := \sum_{t=1}^{T} w_{jt} l_t\left(y_t|y^{t-1}\right) \qquad j \in \{1, ..., T\} \tag{3}$$

where $l_t\left(y_t|y^{t-1}\right)$ is the conditional log-density for observation t and the weights w_{jt} are computed using a kernel function and normalised to sum to one:

$$w_{jt} = \tilde{w}_{jt}/\sum\nolimits_{t=1}^{T} \tilde{w}_{jt}, \quad \tilde{w}_{jt} = K\left(\frac{j-t}{H}\right) \quad \text{for} \quad j, t \in \{1, ..., T\}. \tag{4}$$

The kernel function K is assumed to be non-negative, continuous and bounded function with domain $\mathbb{R}$. The bandwidth parameter H satisfies $H \to \infty$ and $H = o(T/\log T)$. For example, the widely used Normal kernel weights are given by

$$\tilde{w}_{jt} = (1/\sqrt{2\pi})\exp((-1/2)((j-t)/H)^2) \quad \text{for} \quad t, j = 1, .., T, \tag{5}$$

while the rolling window procedure results as a special case of the choice of a flat kernel weights: $w_{jt} = \mathbb{I}(\,|t-j| \leq H)$. For further discussion of the advantages of exponential kernels over the flat kernel for introducing time variation, refer to Giraitis et al. (2014) and Giraitis, Kapetanios and Price (2013). In this setup, Giraitis et al. (2016) show that, under regularity conditions, $\hat{\theta}_j$ is an

$H^{1/2}+(T/H)^{1/2}$- consistent estimator of θ_j for all $j=[\tau T]$, $0<\tau<1$. Furthermore, defining

$$\widehat{\Sigma}_j := \frac{1}{\varkappa_{jT}}\left(-\frac{\partial^2 \ell_j(\hat{\theta}_j)}{\partial\theta\partial\theta'}\right)^{-1}, \qquad \varkappa_{jT} := \left(\sum\nolimits_{t=1}^{T} w_{jt}^2\right)^{-1}, \tag{6}$$

$a.s.$ positive definiteness of $\widehat{\Sigma}_j$ and the bandwidth rate $H=o(T^{1/2})$ are sufficient for asymptotic normality of $\hat{\theta}_j$:

$$\widehat{\Sigma}_j^{-1/2}(\hat{\theta}_j-\theta_j^0) \rightarrow_d \mathcal{N}(0,I) \quad \text{as } T\rightarrow\infty$$

for all $j=[\tau T]$, $0<\tau<1$, with $\mathcal{N}(0,I)$ denoting the multivariate standard normal distribution.

The aim of this Chapter is to provide a Bayesian treatment to this estimation problem. Note that, while $\hat{\theta}_j$ is a consistent extremum estimator, it is not a maximum likelihood estimator because $\sum_{t=1}^{T} w_{jt} l_t\left(y_t|y^{t-1}\right)$ is not a log-likelihood function. Let $\pi_j(\theta_j)$ denote a prior density for θ_j at time j, assumed to be strictly positive and continuous over a compact parameter space Θ. By combining this prior density with the objective function of Giraitis et al. (2016) in (3), we obtain

$$p_j(\theta_j) = \frac{\pi_j(\theta_j)\exp(\ell_j(\theta_j))}{\int_\Theta \pi_j(\theta)\exp(\ell_j(\theta))d\theta}. \tag{7}$$

$p_j(\theta_j)$ is a proper density for θ_j and can be interpreted as an update to the prior beliefs about θ_j after observing the data. In this sense, $p_j(\theta_j)$ can be referred to as a quasi-Bayesian (or Laplace type) posterior density based on statistical learning. Properties of such quasi-posterior densities and the resulting estimators[1] have been studied by Chernozhukov and Hong (2003) and Tian, Liu and Wei (2007). In particular, Chernozhukov and Hong (2003) give conditions under which $p_j(\theta_j)$ can be asymptotically approximated (in total variation of moments norm) by a normal density. As explained in Chernozhukov and Hong (2003), in order for $p_j(\theta_j)$ to be a valid approximation of a limit distribution for θ_j suitable for confidence interval construction, a generalised information matrix equality must apply for an objective function ψ_j that produces an extremum estimator $\hat{\theta}_j$:

$$\lim_{T\rightarrow\infty} Var\left(\hat{\theta}_j\right) = \lim_{T\rightarrow\infty}\left[-\mathbb{E}\mathcal{H}\left(\psi_j\left(\theta_j^0\right)\right)\right]^{-1} \tag{8}$$

where $\mathcal{H}(f)=\partial^2 f/\partial\theta\partial\theta'$ denotes the Hessian matrix of a function $f(\theta)$. Since this asymptotic equivalence does not hold for the local likelihood function ℓ_j of Giraitis et al. (2016), we propose a

[1] A quasi-Bayesian estimator can be defined as the minimiser of some expected loss function, for example for the quadratic loss function, the minimiser is the posterior mean $\int_\Theta \theta_j p_j(\theta_j|Y)d\theta_j$.

modification that re-scales ℓ_j by employing a different weighting scheme:

$$\varphi_j(\theta_j) := \sum_{t=1}^{T} \vartheta_{jt} l_t \left(y_t | y^{t-1}\right) \quad \vartheta_{jt} := \varkappa_{jT} w_{jt} \text{ for } j, t \in \{1, ..., T\} \tag{9}$$

where l_t is the log-likelihood function that appears in (3), and w_{jt} and $\varkappa_{jT}$ are defined in (4) and (6) respectively. It is clear that the objective functions $\varphi_j(\theta_j)$ and $\ell_j(\theta_j)$ give rise to the same maximiser $\hat{\theta}_j$. However, unlike the case $\psi_j = \ell_j$, the generalised information matrix equality (8) holds for the objective function $\psi_j = \varphi_j$, thereby producing a valid posterior variance. The derivation of the asymptotic results based on objective function (9) is summarised by the following proposition.

Proposition 1. *Let Θ be the parameter space of θ_j, and define $J_j(\theta_j^0) := -\frac{1}{\varkappa_{jT}} \mathbb{E}\mathcal{H}\left(\varphi_j\left(\theta_j^0\right)\right)$. Denote the probability density of the normal distribution $\mathcal{N}\left(0, J_j(\theta_j^0)^{-1}\right)$ by*

$$p_\infty(x) := \sqrt{\frac{\left|J_j(\theta_j^0)\right|}{(2\pi)^{\dim(\theta_j^0)}}} \exp\left\{-\frac{1}{2} x' J_j(\theta_j^0) x\right\}.$$

Consider the following family of random vectors:

$$\mathbb{H}_j := \left\{h_j \equiv \sqrt{\varkappa_{jT}}\left(\theta_j - \theta_j^0\right) - J_j(\theta_j^0)^{-1} \nabla\varphi_j(\theta_j^0)/\sqrt{\varkappa_{jT}} : \theta_j \in \Theta\right\}.$$

Then, for all $j = [\tau T]$, $0 < \tau < 1$, the quasi-posterior density *$p_j(h_j)$ of the transformation h_j, based on the objective function (9), for any random vector $h_j \in \mathbb{H}_j$, is asymptotically approximated by $p_\infty(h_j)$ in total variation of moments norm:*

$$\int_{\mathbb{H}_{Tj}} (1 + \|h\|^\alpha) \left|p_j(h) - p_\infty(h)\right| dh \to_p 0 \quad \text{as } T \to \infty$$

for all $\alpha > 0$.

Remarks

1. The assumptions and proof can be found in Appendix 7.1.1. Assumptions 2 and 3 are high level assumptions and require verification on a case-to-case basis[2].

[2] In a more recent version of this work, we provide verification of Assumptions 2 and 3 for the linear Gaussian local likelihood function.

2. By consistency of $\hat{\theta}_j$, $\frac{1}{\varkappa_{jT}}\nabla\varphi_j(\theta_j^0) \to_p 0$ as $T \to \infty$. Proposition 1 therefore implies that when the sample size T is large, draws from $p_j(\theta_j)$ have the following approximate distribution:

$$\theta_j^{sim} \sim \mathcal{N}\left(\theta_j^0, \left[\varkappa_{jT} J_j(\theta_j^0)\right]^{-1}\right).$$

Since $\varkappa_{jT} \to \infty$ for all j, we conclude that the mean and median of the quasi-posterior distribution $p_j(\theta_j)$ are also consistent. Moreover, confidence intervals constructed using the variance of $p_j(\theta_j)$ provide asymptotically valid inference because the quasi-posterior has the same asymptotic variance as the extremum estimator of Giraitis et al. (2016).

3. The violation of the information equality (8) by the objective function ℓ_j of Giraitis et al. (2016) is irrelevant for the frequentist estimators studied in that paper; it only becomes an issue after augmentation of the local likelihood ℓ_j by a prior density. To provide an intuition of why normalisation of the weights by $\varkappa_{jT}$ is required in (9), note that, if the weights w_{jt} satisfy $\sum_{t=1}^{T} w_{jt} = 1$ as in Giraitis et al. (2016), the local likelihood would not dominate the prior as $T \to \infty$. For the prior to vanish, we require that the weights w_{jt} sum to something that increases with the sample size; and in order to obtain the same rate of convergence as in Giraitis et al. (2016), the weights ϑ_{jt} need to sum to $\varkappa_{jT}$.

2.3 Linear Gaussian Univariate Setting

In this Section we derive closed form Normal-Gamma expressions for the quasi-posterior distribution of a linear regression model with time varying parameters. Let

$$y_t = \beta_{0t} + x_{1t}\beta_{1t} + ...x_{kt}\beta_{kt} + \varepsilon_t, \quad \text{where} \quad \varepsilon_t \sim NID(0, \sigma_t^2)$$

and stacking x_t in a $1 \times (k+1)$ vector of exogenous fixed regressors, $x_t := (1, x_{1t}, ..., x_{kt})$ and β_t as a $(k+1) \times 1$ vector of time varying parameters, $\beta_t = (\beta_{0t}, \beta_{1t}, ..., \beta_{kt})'$, we can re-write the model as

$$y_t = x_t\beta_t + \varepsilon_t.$$

ε_t are independent normally distributed mean zero disturbances with a variance σ_t^2, also indexed by time. For convenience, we will work with the precision, denoted by λ_t, and given by the inverse of the variance parameter, $\lambda_t := \sigma_t^{-2}$.

Employing the proposed normalisation of the weights in (9) and following the frequentist approach of Giraitis et al. (2016), the weighted likelihood of the sample $Y := (y_1, ..., y_T)'$ at each point in time j is given by

$$L_j(Y|\beta_j, \lambda_j, X) = (2\pi)^{-S_j/2}\lambda_j^{Sj/2} \exp\left\{-\frac{\lambda_j}{2}\sum_{t=1}^{T}\vartheta_{jt}(y_t - x_t\beta_j)^2\right\}$$

where $S_j = \sum_{t=1}^{T}\vartheta_{jt} = \varkappa_{jT}$ and $\varkappa_{jT}$ is defined in (6). One can equivalently write the local likelihood in a more compact form as

$$L_j(Y|\beta_j, \lambda_j, X) = (2\pi)^{-S_j/2}\lambda_j^{Wj/2} \exp\left\{-\frac{\lambda_j}{2}(Y - X\beta_j)'D_j(Y - X\beta_j)\right\} \tag{10}$$

where $Y := (y_1, ..., y_T)'$ is a $T\times 1$ vector, $X = (x_1', ..., x_T')'$ is a $T\times k$ matrix and D_j is a $T\times T$ diagonal weighting matrix, containing the kernel weights in its main diagnal, $D_j = diag(\vartheta_{j1}, ..., \vartheta_{jT})$.

Next, assume that β_j and λ_j have Normal-Gamma prior distribution for $j \in \{1, ..., T\}$:

$$\beta_j|\lambda_j \sim \mathcal{N}\left(\beta_{0j}, (\lambda_j\kappa_{0j})^{-1}\right), \quad \lambda_j \sim Ga(\alpha_{0j}, \gamma_{0j}) \tag{11}$$

where β_{0j} is a $k \times 1$ vector of prior means, κ_{0j} is a $k \times k$ positive definite symmetric matrix, and α_{0j} and γ_{0j} are the shape and scale parameters of the Gamma distribution respectively.

Proposition 2. *Combining the weighted likelihood L_j in (10) with the prior in (11), β_j and λ_j have Normal-Gamma quasi-posterior distribution for $j = \{1, ..., T\}$:*

$$\begin{aligned} \beta_j|\lambda_j, X, Y &\sim \mathcal{N}\left(\widetilde{\beta}_j, (\lambda_j\widetilde{\kappa}_j)^{-1}\right), \\ \lambda_j &\sim Ga(\widetilde{\alpha}_j, \widetilde{\gamma}_j), \end{aligned} \tag{12}$$

with posterior parameters:

$$\widetilde{\beta}_j = \widetilde{\kappa}_j^{-1}(X'D_jX\hat{\beta}_j + \kappa_{0j}\beta_{0j})$$

$$\widetilde{\kappa}_j = \kappa_{0j} + X'D_jX, \quad \widetilde{\alpha}_j = \alpha_{0j} + S_j/2$$

$$\widetilde{\gamma}_j = \gamma_{0j} + \frac{1}{2}\left(Y'D_jY - \widetilde{\beta}_j'\widetilde{\kappa}_j\widetilde{\beta}_j + \beta_{0j}'\kappa_{0j}\beta_{0j}\right)^3$$

where

$$\hat{\beta}_j = (X'D_jX)^{-1}X'D_jy$$

is the local likelihood estimator for β_j.

Proposition 3. *The marginal distribution of* β_j *is given by a multivariate non-standardised t-distribution with* $2\widetilde{\alpha}_j$ *degrees of freedom:*

$$\beta_j|Y, X \sim T_{\widetilde{\alpha}_j}\left(\tilde{\beta}_j, \frac{\tilde{\gamma}_j}{\tilde{\alpha}_j}\tilde{\kappa}_j^{-1}\right)$$

where $\widetilde{\alpha}_j$, $\widetilde{\beta}_j$, $\widetilde{\gamma}_j$ *and* $\widetilde{\kappa}_j$ *are defined in Proposition 2.*

2.4 Linear Gaussian Multivariate Setting

In this Section we derive closed form Normal-Wishart expressions for the quasi-posterior distribution of a VAR(k) model with time varying parameters.

Suppose that we have an $M \times 1$ dimensional vector y_t generated by a time varying parameter (TVP) VAR model of lag order k. Then, y_t can be written as:

$$y_t = B_{0t} + \sum\nolimits_{p=1}^{k} B_{pt} y_{t-p} + \varepsilon_t \tag{13}$$

where B_{0t} is an $M \times 1$ vector of time varying intercepts, and B_{pt} is an $M \times M$ matrix of time varying autoregressive coefficients for lag $p = 1, ..., k$. The error term, ε_t, is an $M \times 1$ vector of normally distributed zero mean random variables, with a positive definite symmetric $M \times M$ contemporaneous covariance matrix R_t^{-1}, which may be varying over time. Then, ε_t can be written as: $\varepsilon_t = R_t^{-1/2}\eta_t$ where $\eta_t \sim NID(0_M, I_M)$. In addition, denote by $x_t := (1, y'_{t-1}, ..., y'_{t-k})$ a $1 \times (Mk+1)$ vector and by $B_t := (B_{0t}, B_{1t}, ..., B_{kt})$ an $M \times (Mk+1)$ matrix. Then, the model (13) can be written as

$$y_t = B_t x'_t + \varepsilon_t$$

After vectorising, one can write

$$y_t = (I_M \otimes x_t)\beta_t + R_t^{-1/2}\eta_t, \tag{14}$$

where $\beta_t := vec(B'_t)$ is an $M(Mk+1) \times 1$ vector for $t = 1, ..., T$.

Following the approach of Giraitis et al. (2016) and employing the proposed normalisation of the weights in (9), the weighted likelihood of the sample $(y_1, ..., y_T)$ for the VAR(k) model (13) at each point in time j is given by

$$L_j(y|\beta_j, R_j, X) = (2\pi)^{-MS_j/2} |R_j|^{S_j/2} e^{-\frac{1}{2}\sum_{t=1}^{T} \vartheta_{jt}(y_t-(I_M\otimes x_t)\beta_j)' R_j (y_t-(I_M\otimes x_t)\beta_j)} \tag{15}$$

where $S_j = \sum_{t=1}^{T} \vartheta_{jt}$ and the kernel weights ϑ_{jt} are defined in (9). Denote by $Y = (y_1, ..., y_T)'$ a $T \times M$ matrix of stacked vectors $y_1', ..., y_T'$ and define $y = vec(Y)$ as a $TM \times 1$ vector. Similarly, define $E = (\varepsilon_1, ..., \varepsilon_T)'$ and $TM \times 1$ vector $\varepsilon = vec(E)$. Let X be a $T \times Mk+1$ matrix defined as $X = (x_1', ..., x_T')'$. Then the weighted likelihood (15) can be written in a more compact form as:

$$L_j(y|\beta_j, R_j, X) \propto |R_j|^{S_j/2} \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta_j)'(R_j \otimes D_j)(y - (I_M \otimes X)\beta_j)\right\} \tag{16}$$

where $D_j := diag(\vartheta_{j1}, ..., \vartheta_{jT})$ for $j \in \{1..., T\}$.

Next, assume that β_j and R_j have Normal-Wishart prior distribution for $j \in \{1, ..., T\}$:

$$\beta_j|R_j \sim \mathcal{N}\left(\beta_{0j}, (R_j \otimes \kappa_{0j})^{-1}\right), \quad R_j \sim W(\alpha_{0j}, \gamma_{0j}) \tag{17}$$

where β_{0j} is a $(Mk+1)M \times 1$ vector of prior means, κ_{0j} is a $(Mk+1) \times (Mk+1)$ positive definite symmetric matrix, α_{0j} is a scalar scale parameter of the Wishart distribution and γ_{0j} is a $M \times M$ positive definite symmetric matrix.

Proposition 4. *Combining the weighted likelihood L_j in (16) with the prior in (17), β_j and R_j have Normal-Wishart quasi-posterior distribution for $j = \{1, ..., T\}$:*

$$\begin{aligned} \beta_j|R_j, X, Y &\sim \mathcal{N}\left(\widetilde{\beta}_j, (R_j \otimes \widetilde{\kappa}_j)^{-1}\right), \\ R_j &\sim W(\widetilde{\alpha}_j, \widetilde{\gamma}_j), \end{aligned} \tag{18}$$

with posterior parameters:

$$\begin{aligned} \widetilde{\beta}_j &= \left(I_M \otimes \widetilde{\kappa}_j^{-1}\right)\left[(I_M \otimes X'D_jX)\hat{\beta}_j + (I_M \otimes \kappa_{0j})\beta_{0j}\right], \\ \widetilde{\kappa}_j &= \kappa_{0j} + X'D_jX, \quad \widetilde{\alpha}_j = \alpha_{0j} + S_j, \\ \widetilde{\gamma}_j &= \gamma_{0j} + Y'D_jY + B_{0j}\kappa_{0j}B_{0j}' - \widetilde{B}_j\widetilde{\kappa}_j\widetilde{B}_j', \end{aligned} \tag{19}$$

where

$$\hat{\beta}_j = (I_M \otimes X'D_jX)^{-1}(I_M \otimes X'D_j)y \tag{20}$$

is the local likelihood estimator for β_j.

The marginal distribution of the parameter vector β_j can be obtained by integrating the joint quasi-posterior distribution $p(\beta_j, R_j|Y, X)$ obtained in Proposition 4 over the $M(M+1)/2$ distinct elements of the symmetric matrix R_j.

Proposition 5. *The marginal distribution of β_j is given by a multivariate non-standardised t-distribution with $\widetilde{\alpha}_j - Mk$ degrees of freedom:*

$$\beta_j|Y, X \sim T_{\widetilde{\alpha}_j - Mk}\left(\widetilde{\beta}_j, \frac{\widetilde{\gamma}_j \otimes \widetilde{\kappa}_j^{-1}}{\widetilde{\alpha}_j - Mk - 2}\right)$$

where $\widetilde{\alpha}_j$, $\widetilde{\beta}_j$, $\widetilde{\gamma}_j$ and $\widetilde{\kappa}_j$ are defined in Proposition 4.

2.5 Conditional quasi-posterior distributions

There are two cases of particular interest to be considered for the model (14): first, when the covariance matrix $R_t^{-1/2}$ is known; second, when the parameter vector β_t is known. The resulting closed form conditional quasi-posterior densities, derived in Propositions 4 and 5 below, are useful for the design of the Gibbs algorithms for estimating model (14) generated by mixtures of fixed and time varying parameters. These Gibbs algorithms are developed in Section 2.8.

Case 1. For known covariance matrix R_t^{-1}, the model (14) can be transformed in the following way:

$$\begin{aligned}
\widetilde{y}_t &= \widetilde{x}_t\beta_t + \eta_t, \quad \eta_t \sim NID(0_M, I_M), \\
\widetilde{y}_t &= R_t^{1/2}y_t, \;\; \widetilde{x}_t = R_t^{1/2}(I_M \otimes x_t).
\end{aligned} \tag{21}$$

Case 2. For known β_t, the model (13) can be written as

$$\varepsilon_t = y_t - (I_M \otimes x_t)\beta_t = R_t^{-1/2}\eta_t, \;\; \eta_t \sim NID(0_M, I_M).$$

Proposition 6. *If a $\mathcal{N}\left(\beta_{0j}, V_{0j}^{-1}\right)$ prior distribution is selected for β_j in Case 1, the quasi-posterior distribution of β_j* is given by $\mathcal{N}\left(\widetilde{\beta}_j, \widetilde{V}_j^{-1}\right)$ *with posterior mean and variance*

$$\widetilde{\beta}_j = \widetilde{V}_j^{-1}\left[\sum\nolimits_{t=1}^{T} \vartheta_{jt}\widetilde{x}_t'\widetilde{y}_t + V_{0j}\beta_{0j}\right], \;\; \widetilde{V}_j = V_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt}\widetilde{x}_t'\widetilde{x}_t \tag{22}$$

for each $j \in \{1, ..., T\}$*, where* $\widetilde{x}_t$ *and* $\widetilde{y}_t$ *are defined in (21).*

Proposition 7. *If a Wishart* $W(\alpha_{0j}, \gamma_{0j})$ *prior distribution is selected for* R_j *in Case 2, the quasi-posterior of* R_j *is also Wishart* $W(\widetilde{\alpha}_j, \widetilde{\gamma}_j)$ *with posterior parameters*

$$\widetilde{\alpha}_j = \alpha_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt}, \quad \widetilde{\gamma}_j = \gamma_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt} \varepsilon_t' \varepsilon_t \tag{23}$$

for each $j \in \{1, ..., T\}$ *.*

2.6 Optimal prior shrinkage and optimal lag

If the VAR dimension is large, it is convenient to have an automatic way to select the prior tightness κ_{0j} and the most widely used approaches are the Minnesota prior proposed by Litterman (1980) and the sum of coefficients prior developed by Doan, Litterman and Sims (1984). These require that the researcher chooses a small number of hyperparameters that control the overall tightness of the prior. The question of optimal (i.e. data-driven) choice of hyperparameters and the related question of choosing optimal lag order for the VAR model has received a lot of attention in the Bayesian literature with papers such as Phillips (1995), Carriero, Kapetanios and Marcellino (2012), Carriero, Clark and Marcellino (2015a), Giannone, Lenza and Primiceri (2015). To this end, Carriero et al. (2012, 2015) maximise the marginal likelihood of the data, given by

$$p(Y) = \int \int p(Y|\beta, R) p(\beta|R) p(R) d\beta dR$$

over a fine grid of points for the overall prior tightness (or the lag order). Giannone et al. (2015) use a hierarchical prior for the overall tightness parameter ζ, $p(\zeta)$ which then delivers a posterior

$$p(\zeta|Y) \propto p(Y|\zeta) p(\zeta).$$

In the Proposition 8 below, we derive a closed form expression for the quasi-marginal likelihood $p(Y)$ in the setup of time varying parameter VAR model. There is one such expression for each point in time, so that if the researcher wants to obtain optimal shrinkage (or lag order) for the estimation of the parameters at some point in time j, she can maximise the analytic expression below at time j with respect to the overall shrinkage hyperparameter or proceed by specifying a hierarchical prior for it.

Proposition 8. The quasi-marginal likelihood of the sample Y is given by

$$\begin{aligned} p_j(Y) &= \int\int p(Y|\beta, R)p(\beta|R)p(R)d\beta dR \\ &= \frac{(2)^{(M-1)S_j/2}\left|\kappa_{0j}\right|^{(Mk+1)/2}\Gamma_M\left(\tilde{\alpha}_j/2\right)\left|\gamma_{0j}\right|^{\alpha_{0j}/2}}{\pi^{S_j/2}\left|\widetilde{\kappa}_j\right|^{-(Mk+1)/2}\Gamma_M\left(\alpha_{0j}/2\right)\left|\widetilde{\gamma}_j\right|^{\widetilde{\alpha}_j/2}} \qquad \text{for } j \in \{1, ..., T\}. \end{aligned}$$

Remarks.

1. The quasi-posteriors derived in Proposition 2 and Proposition 4 share similarity with the Normal-Gamma and Normal-Wishart conjugate posterior results in the Bayesian VAR literature respectively. However, the QBLL approach is augmented to include the kernel weights ϑ_{jt} and provides posterior distributions for each point in time j, thereby accommodating time variation in the parameters. Note that the result in Proposition 1 applies to a system of a general form; however, since the focus of this Chapter is linear Gaussian models, we restricts attention on VAR systems.

2. While the model in (13) is of parametric form, the time variation in the parameters is dealt with nonparametrically: the sequence of parameters needs only satisfy one of the "slow drift" conditions (1) or (2), which encompass a wide class of processes without the need to impose a specific modelling restrictions on the parameters.

3. The QBLL approach provides a framework for accommodating prior beliefs on the presence of time variation in the parameters, since the prior parameters are allowed to vary over time.

4. Quasi-Bayesian point estimators for β_j and R_j^{-1} can be obtained by using the means of the quasi-posterior distribution in Proposition 4. For β_j, the posterior mean is given by $\widetilde{\beta}_j$ in (19). For R_j^{-1}, the Wishart property of R_j implies an inverse Wishart $IW(\tilde{\alpha}_j, \tilde{\gamma}_j)$ distribution for R_j^{-1}, with posterior mean given by $\tilde{\gamma}_j/(\tilde{\alpha}_j - M - 1)$, in the notation of Proposition 4.

5. In view of the modification of the kernel weights in (9), the quasi-Bayesian estimators in Remark 4 are asymptotically equivalent to the local likelihood estimator of Giraitis et al. (2016) and thereby inherit the latter's consistency and asymptotic normality properties. In particular, $\widetilde{\beta}_j$ is a weighted average of the prior mean, β_{0j}, and the local likelihood estimator in (20) with relative weights given by $I_M \otimes (\kappa_{0j} + X'D_jX)^{-1}\kappa_{0j}$ and $I_M \otimes (\kappa_{0j} + X'D_jX)^{-1}X'D_jX$ respectively. The former expression depends positively on the prior tightness κ_{0j}^{-1} and converges to zero as $T \to \infty$; the latter expression converges to 1. Convergence of both expressions (and

hence asymptotic negligibility of the priors) applies because the modified kernel weights ϑ_{jt} ensure that $\text{plim}_{T\to\infty} X'D_jX = \infty$. Similarly, the posterior mean for R_j^{-1} is asymptotically equivalent to the local likelihood estimator of Giraitis et al. (2016).

6. Despite the asymptotic equivalence of the quasi-Bayesian and local likelihood estimators discussed in Remark 5, the prior distribution may have a considerable small sample effect: a suitable choice of prior can improve the finite sample performance of the frequentist estimator and can help avoid over-fitting in VAR systems. Moreover, the prior plays an essential role in large-dimensional VAR systems, where the local likelihood estimator may be undefined due to lack of invertibility of $X'D_jX$. In such cases, the presence of the prior ensures feasibility of the QBLL estimator.

2.7 Discussion

We provide a discussion of the relative merits of the proposed QBLL approach compared to widely used state space approaches to the estimation of the time varying parameters β_j and R_j^{-1}. The most serious limitation of state space methodology is the inability to accommodate large dimensional VAR systems, and we provide an additional comparison of our QBLL approach to alternative methods that are able to deal with large dimensions.

2.7.1 Comparison to state space models

Writing a VAR model in state space requires adding the drifting parameters to a state vector of latent variables. This involves the specification of a stochastic process for modelling these unobserved parameters, with a random walk process being the most common assumption in the literature (see for example, Cogley and Sargent (2002, 2005), Primiceri (2005), Mumtaz and Surico (2009), Cogley et al. (2010) and Clark (2012)). This assumption is convenient as it is the simplest way to induce persistence in the parameters and, at the same time, considerably reduces the number of additional coefficients added to each state equation. However, inaccurate specification of the state equation as a random walk invalidates inference, even asymptotically. The finite sample distortions resulting from such misspecification can be as severe as coverage rates of 30% for a nominal of 95%, as demonstrated in the Monte Carlo exercise in Section 2.9. Similar evidence for a univariate model is presented in Appendix 7.1.11. State space models also require additional assumptions about how the innovations in the newly added state variables are correlated across equations.

Although state space models make use of Bayesian tools such as MCMC algorithms, it is arguable that the drifting parameters no longer receive a Bayesian treatment. While prior distributions are specified for the initial values of the filter and for the additional coefficients guiding the processes, the actual drifting parameters are now latent variables. Consequently, some of the benefits of Bayesian estimation are forgone and state space models can run into issues such as non-stationary draws from the time varying autoregressive matrices or values that make little economic sense. These difficulties can be overcome if prior distributions for the drifting parameters can be directly specified at each point in time, as is the case with the proposed QBLL estimators.

Another issue that arises when modelling the drifting covariance matrix R_j^{-1} in a state space setup is the potential loss of the positive definite property of this matrix, if the elements of R_j^{-1} are modelled as drifting stochastic processes. To address this issue, Cogley and Sargent (2005), Primiceri (2005) and Cogley et al. (2010) diagonalise the covariance matrix using Cholesky decomposition and then assume that the diagonal elements follow a random walk in logarithms, thus ensuring they remain positive. This requires additional modelling assumptions relating to the stochastic properties of the elements of lower triangular matrix in the Cholesky decomposition. Diagonalisation of the covariance matrix is useful for conducting structural analysis and Canova and Perez Forero (2015) shows how to estimate time varying identifying restrictions that are not necessarily just-identified and recursive. However, in models where structural shock analysis is not required, diagonalisation has only been employed as means to facilitate estimation of the drifting volatilities (Cogley and Sargent (2005), Clark (2012)). The QBLL approach permits direct estimation of R_j^{-1}, which has an inverted-Wishart posterior density and hence remains symmetric and positive definite at each point in time, thus making diagonalisation redundant. When applied to structural shock analysis, the QBLL estimator is informative on the presence of time variation in the identifying restrictions, this information being an outcome of the estimation procedure rather than a maintained assumption.

Perhaps the most important limitation to the practical application of state space methodology is their inability to accommodate large dimensional models. The size and complexity of the state space increases substantially with the VAR dimension, since an extra state equation is required for each parameter as well as an additional shock and additional coefficients guiding the process[4]. As a result, state space models suffer from dimensionality problems and their application to the

[4]For instance, in a state space setting, a four variables TVP VAR (2) with stochastic volatility requires the addition of 40 state equations.

estimation of TVP-VAR models is limited to a model of three to four variables. Additional complexity in the estimation procedure of state space models arises from numerical approximation of the joint posterior density conducted by MCMC algorithms. On the other hand, the proposed QBLL methodology admits a closed form joint posterior density, allowing estimation of VAR models of size of over 80 variables[5], as demonstrated in Section 2.10.2.

2.7.2 Comparison to alternative methods for large TVP-VARs

To address the dimensionality problems of state space models, Koop and Korobilis (2013) employ a 'forgetting factor' approach, which involves discounting past observations in the Kalman filter prediction step. Their approach provides a remedy to the dimensionality problem but remains susceptible to the modelling limitations of state space methods discussed in the previous Section; e.g. lack of robustness to misspecification of the state equations.

The paper of Kapetanios, Marcellino and Venditti (2015) also proposes a solution to the dimensionality problem, employing a procedure based on the frequentist kernel estimator of Giraitis et al. (2016). They propose a stochastic constraint estimator and a ridge estimator in order to achieve shrinkage of their VAR parameters and avoid over-fitting. Their stochastic constraint estimator coincides with the mean of the QBLL posterior in equation (19), while their ridge estimator is a special case of (19). The treatment of the VAR parameters in Kapetanios et al. (2015) is frequentist and does not allow for: (i) posterior distributions varying over time; (ii) a mechanism for taking into account prior beliefs about the volatilities. The QBLL approach produces a closed form expression for the joint posterior density in (18) and can incorporate prior beliefs about R_j^{-1} in the form of the parameters α_{0j} and γ_{0j}. An additional advantage of the methodology of this Chapter, is that Bayesian analysis provides a framework for the construction of MCMC algorithms which enables the estimation of mixtures of time varying and time invariant parameters.

2.8 A Gibbs sampling approach

There are cases when the researcher would like to consider the model (13) outlined in Section 2.5 with time invariant autoregressive coefficients β and time varying variances R_t^{-1}. Such models have been proposed by Primiceri (2005) and Sims and Zha (2006) have become popular for being able to fit well macroeconomic data when the sample considered contains periods characterised

[5] Such a VAR(1) model with 80 variables with time variation in the parameters and the covariance matrix would require the use of 9720 state equations.

by different volatility; for example, data containing the high volatility periods of the oil crises followed by the low volatility periods of the Great Moderation. Alternatively, the researcher might want to keep the variances of the errors fixed over time and allow for time varying autoregressive coefficients, as in Cogley and Sargent (2002). This approach also has merits: an AR model with drifting autoregressive coefficients is a flexible framework that can fit a range of nonlinearities. Such models are popular and widely used in other disciplines such as image and signal processing (see for example, Kitagawa (1996), Abramovich, Spencer and Turley (2007)). Another approach useful in some applications assumes most parameters to be fixed but one or two, interesting from economic perspective, are allowed to vary over time.

It is therefore desirable to develop a framework of inference that encompasses all the above scenaria by allowing for mixtures of time varying and time invariant parameters. To this end, we utilise the Bayesian theory developed in Section 2.5 in order to design several Gibbs algorithms that can produce approximate joint posterior distributions for mixtures of time varying and time invariant parameters. Such cases are not covered by the estimators of Giraitis et al. (2014, 2016) and are only covered by state space models at the cost of possible state equation misspecification and VAR dimensionality restrictions, as discussed in Section 2.7. The algorithms proposed below perform well with different parameter processes and allow the researcher to remain agnostic about the parameters' data generating process.

2.8.1 Homoscedastic BVAR model with time varying parameters

Consider an M-dimensional process y_t defined by a VAR model with time-varying autoregressive parameters and time invariant volatility: $y_t = (I_M \otimes x_t)\beta_t + R^{-1/2}\eta_t,\ \ \eta \sim \mathcal{N}(0_M, I_M)$, where x_t is a $1 \times Mk + 1$ vector process defined in (13) and β_t is $M(Mk+1) \times 1$ vector of time varying coefficients.

A Gibbs algorithm that can sample from the joint quasi-posterior of β_t and R^{-1} can be constructed in the following way. Conditional on a draw R, the model can be redefined as

$$\widetilde{y}_t = R^{1/2} y_t = \widetilde{x}_t \beta_t + \eta_t \tag{24}$$

where $\widetilde{x}_t := R^{1/2}(I_M \otimes x_t)$. Assuming a normal prior $\mathcal{N}(\beta_{0j}, V_{0j})$ for the parameter process β_j, the quasi-posterior distribution of β_j, established in Proposition 6, is also normal $\mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$ at each point in time j with parameters given in (22). On the other hand, conditional on a draw from the history of β_j, $\beta_{1:T} = (\beta_1, ..., \beta_T)$, the model (24) can be written as $\widetilde{\varepsilon}_t = y_t - (I_M \otimes X_t)\beta_t = R^{-1/2}\eta_t$. Then, assuming a Wishart prior distribution $W(\alpha_0, \gamma_0)$ for R^{-1} and stacking $\widetilde{E} = (\widetilde{\varepsilon}_1, \widetilde{\varepsilon}_2, ..., \widetilde{\varepsilon}_{tT})'$, the conditional posterior of R^{-1} can be written as $R^{-1}|X, Y, \beta_{1:T} \sim W(\widetilde{\alpha}, \widetilde{\gamma})$, where

$$\widetilde{\alpha} = \alpha_0 + T, \quad \widetilde{\gamma} = \gamma_0 + \widetilde{E}'\widetilde{E}. \tag{25}$$

The conditional posterior of R^{-1} is of standard form and the conditional quasi-posterior distribution of β_j, characterised in Proposition 6, can be easily drawn from; hence, the estimation of the model in (24) permits the use of a Gibbs algorithm with the following steps.

Algorithm 1. **Step 1.** Initialise the algorithm with a guess, $R^{-1,0}$.

Then for $i = 1, ..., N$, iterate between steps 2 and 3.

Step 2. Draw $\beta_j^i|Y, X, R^{-1,i-1}$ from $\mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$ with posterior parameters defined in (22) for each point in time $j \in \{1, ..., T\}$.

Step 3. Draw $R^{-1,i}|X, Y, \beta_{1:T}^i$ from $W(\widetilde{\alpha}, \widetilde{\gamma})$ with posterior parameters defined in (25).

Standard Markov chain Monte Carlo (MCMC) results apply as follows. Since the form of the quasi-posterior distributions has been established and since the steps above constitute a Markov chain, iterating between step 2 and 3 results into convergence of the distribution of the chain to its stationary distribution and hence, draws from the algorithm (after discarding a fair proportion of initial draws) can be used for an approximation of the joint quasi-posterior distribution of β_j and R^{-1}.

2.8.2 Heteroscedastic BVAR model with fixed parameters

Consider an M-dimensional process y_t defined by the model

$$y_t = (I_M \otimes x_t)\beta + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, R_t^{-1}) \tag{26}$$

where x_t is a $1 \times Mk + 1$ vector process defined in (13) and β is $M(Mk + 1) \times 1$ vector of time invariant coefficients. The error term ε_t has a time varying variance covariance matrix R_t^{-1}. Conditional on observing a draw of $R_{1:T}^{-1}$ from the history $(R_1^{-1}, ..., R_T^{-1})$, the model reduces to a GLS problem with known time varying covariance matrix. Pre-multiplying equation (26) with $R_t^{1/2}$, we obtain $R_t^{1/2} y_t = R_t^{1/2}(I_M \otimes x_t)\beta + \eta_t$, $\eta_t \sim \mathcal{N}(0_M, I_M)$, and after defining $\widehat{y}_t := R_t^{1/2} y_t$ and $\widehat{x}_t := R_t^{1/2}(I_M \otimes x_t)$, $\widehat{y}_t$ follows a homoscedastic model $\widehat{y}_t = \widehat{x}_t\beta + \eta_t$. Assuming a normal prior

$\mathcal{N}(\beta_0, V_0)$ for β, the posterior distribution is also normal $\mathcal{N}(\widetilde{\beta}, \widetilde{V})$, with parameters

$$\widetilde{V} = \left(\sum_{t=1}^{T} \widetilde{x}_t' \widetilde{x}_t + V_0^{-1} \right)^{-1} , \quad \widetilde{\beta} = \widetilde{V} \left(\sum_{t=1}^{T} \widetilde{x}_t' \widetilde{y}_t + V_0^{-1} \beta_0 \right). \tag{27}$$

Conditional on observing the coefficients β, the model simplifies to $\varepsilon_t = y_t - (I_M \otimes X_t)\beta$ where $\varepsilon_t \sim \mathcal{N}(0_M, R_t^{-1})$ and is also observed. Then, by assuming a Wishart prior $W(\alpha_{0j}, \gamma_{0j})$ for R_j^{-1} at each point in time $j \in \{1, ..., T\}$, Proposition 7 implies that the conditional quasi-posterior of R_j^{-1} at each point in time j is also Wishart $W(\widetilde{\alpha}_j, \widetilde{\gamma}_j)$, with parameters given in (23).

Since, the conditional posterior of β is of standard form and the quasi-posterior distribution of R_j^{-1} has been characterised in Section 2.5, a Gibbs algorithm can be constructed in order to recursively draw from the conditional posterior distributions of β and $R_{1:T}^{-1}$ to approximate their joint posterior distribution. The algorithm consists of three steps.

Algorithm 2. **Step 1**. Initialise the algorithm with β^0.

For $i = 1, ..., N$ iterate between steps 2 and 3.

Step 2. For each $j \in \{1, ..., T\}$ draw $R_j^{-1,i} | X, Y, \beta^{i-1}$ from $W(\widetilde{\alpha}_j, \widetilde{\gamma}_j)$ with posterior parameters defined in (23).

Step 3. Draw $\beta^i | Y, X, R_{1:T}^{-1,i}$ from $\mathcal{N}(\widetilde{\beta}, \widetilde{V})$ with posterior parameters defined in (27).

Note that it is easy to extend the models (24) and (26) respectively to include: i) a set of regressors which enter (24) with time invariant coefficients ζ, in addition to the regressors X_t which enter with time varying coefficients β_t; ii) a set of regressors which enter (26) with time varying coefficients ζ_t in addition to the regressors x_t which enter with invariant coefficients, β. For space consideration, the corresponding algorithms are outlined in Appendix 7.1.9.

2.8.3 Time varying structural BVAR model

Algorithms 1 and 2 involve reduced-form VAR models. In macroeconomic applications, structural VARs (SVARs) are often applied to analyse macroeconomic shocks and transmission of macroeconomic shocks to key variables. SVAR models orthogonalise the VAR residuals so that the resulting 'structural' shocks have a diagonal covariance matrix. The simplest such orthogonalisation of the $M \times M$ covariance matrix R_t^{-1} considered in the literature involves a Cholesky decomposition. In particular, consider the model:

$$y_t = (I_M \otimes X_t)\beta_t + \varepsilon_t, \quad var(\varepsilon_t) = R_t^{-1} = A^{-1}\Omega_t^{-1}A^{-1\prime} \tag{28}$$

where Ω_t^{-1} is diagonal matrix with elements ω_{it}, varying over time, and A is a lower triangular matrix with ones on its main diagonal. Providing a Bayesian treatment for the $\frac{1}{2}M(M-1)$ non-zero elements of A is straightforward. As shown in Cogley and Sargent (2002), the VAR model can be written as

$$A(y_t - (I_M \otimes X_t)\beta_t) := Ay_t^* = \Omega_t^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}(0, I_M). \tag{29}$$

As argued in Primiceri (2005), the researcher might also want to allow the orthogonalisation scheme of the structural shocks to be time-varying. In Primiceri (2005), this involves adding the non zero or one elements of A_t to the state vector and specifying stochastic processes for them (in his application, these follow a random walk process). Below, we present an alternative algorithm that can provide estimates over time of A_t, by employing the nonparametric kernel-based QBLL method, which does not require specifying parameter process.

Conditioning on β_t and Ω_t, the model in (28) with drifting lower triangular matrix A_t simplifies to a set of $k-1$ equations in (29), $i = 2, ..., k$, with equation i having y_{it}^* as a dependent variable and $-y_{jt}^*$ as regressor:

$$\underbrace{\omega_{it}^{-1}y_{it}^*}_{\widehat{Y}_{it}} = \underbrace{-\omega_{it}^{-1}[y_{1t}^*, ..., y_{i-1,t}^*]}_{\widehat{X}_{it}} \begin{bmatrix} a_{i1,t} \\ \vdots \\ a_{i,i-1,t} \end{bmatrix} + \eta_{it}.$$

Assuming a normal prior $\mathcal{N}(a_{0,ij}, \lambda_{0,ij})$ for a_{ij}, the coefficients of the i^{th} row of matrix A_j at each point in time $j \in \{1, ..., T\}$, the quasi-posterior distribution of a_{ij} is also normal $N\left(\widetilde{a}_j, \widetilde{V}_j\right)$,with parameters

$$\widetilde{V}_j = \left(\lambda_{0,ij}^{-1} + \sum_{t=1}^{T} \vartheta_{jt}\widehat{X}_t'\widehat{X}_t\right)^{-1}, \quad \widetilde{a}_j = \widetilde{V}_j^{-1}\left(\sum_{t=1}^{T} \vartheta_{jt}\widehat{X}_{it}'\widehat{Y}_{it} + \lambda_{0,ij}^{-1}a_{0,ij}\right). \tag{30}$$

It is straightforward to build this step into one of the previously outlined Gibbs algorithms (with or without time variation in the VAR's autoregressive parameters or in the volatility) in order to draw β_t, Ω_t and A_t from their conditionally conjugate quasi-(or standard) posterior distributions in order to approximate their joint quasi-posterior distribution. The kernel approach can be extended to the estimation of not necessarily just-identified or recursive time varying identifying restrictions; we leave this question for future research.

2.9 Monte Carlo

In this Section, we design a Monte Carlo exercise to study the finite sample properties of the QBLL estimators introduced in this Chapter and how they compare to the alternative state space approach. We limit our attention to a two dimensional VAR(1) model with no intercept term (Appendix 7.1.11 contains some earlier Monte Carlo results with a univariate models and also compares the QBLL approach with non-linear particle filter). We simulate data using four different data generating processes (DGPs) and estimate the models we want to compare based on these simulated samples. Sample sizes of 100, 500 and 1000 are considered and due to computational considerations, 300 replications for each DGP and sample size are simulated. In each case, we compute the bias and the root mean square error (RMSE) of the estimators. For simplicity, the average bias and the root of the MSE over time are reported, summarised averaging over the autoregressive and covariance parameters respectively. We also report the 95% coverage rates for the parameter estimates, computed as the proportion of time that the true parameter finds itself in the 95% posterior confidence intervals implied by the different models. For the QBLL approach, we use the means of the quasi-posterior density as point estimates, the normal kernel in (5) for the kernel weights with bandwidth of $T^{0.5}$ and a Minnesota type prior[6] with loose overall shrinkage $\lambda = 1$. Our choice of bandwidth $H = T^{0.5}$ is motivated by the optimal bandwidth choice used for inference in time varying random coefficient models, see Giraitis et al. (2014). For all state space models in the Monte Carlo exercise, we use the algorithm outlined in Cogley and Sargent (2005) with 3,000 Gibbs draws[7], modified to allow for time varying Cholesky lower triangular matrix, whose elements are modelled as a random walk, as in Primiceri (2005). The priors and initial values of the state space models are set using the initial 10% of the observations as a presample.

[6] The priors for the autoregressive parameters are as in Bańbura, Giannone and Reichlin (2010) with $\lambda = 1$ and the Wishart prior parameters for the volatilities are as in Kadiyala and Karlsson (1997).

[7] From which the first 1,000 have been discarded.

DGP I. The first DGP we consider is a model with fixed parameters and fixed volatility:

$$\begin{aligned} y_t &= By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Psi), \quad y_0 = 0, \\ B &= \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.6 \end{bmatrix} \text{ and } \Psi = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}. \end{aligned} \tag{31}$$

DGP I								
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
T=100	QBLL_TVP_TVV	-0.0307	0.1698	0.9326	-0.0105	0.0605	0.9292
	SS_TVP_TVV	-0.0207	0.1275	0.9526	0.0002	0.0574	0.9032
T=500	QBLL_TVP_TVV	-0.0134	0.1033	0.9457	-0.0030	0.0410	0.9383
	SS_TVP_TVV	-0.0072	0.0779	0.9163	-0.0026	0.0347	0.8592
T=1000	QBLL_TVP_TVV	-0.0126	0.0874	0.9399	-0.0051	0.0341	0.9424
	SS_TVP_TVV	-0.0081	0.0677	0.8957	-0.0077	0.0318	0.8197

Table 1. Bias, RMSEs and coverage rates of models based on DGP I for 100, 500 and 1000 observations respectively.

Our motivation is to assess how well models featuring time varying parameters can fit a simple time invariant model. Table 1 presents the bias, RMSE and coverage rates for the QBLL estimator featuring time varying parameters and volatility (QBLL_TVP_TVV) and a state space model, with drifting parameters and volatilities, referred to as SS_TVP_TVV. It is clear from Table 1 that both models deliver valid inference. For small samples, the state space approach has smaller RMSEs, but when the sample size increases, both models deliver similar performance. With the increase of the sample size, the coverage rate of the state space model for both the autoregressive parameters and the variance deteriorates, and when the sample size is 1000, the coverage rate is considerably below the nominal 95%, implying the presence of type I errors. On the other hand, the confidence intervals implied by the QBLL approach remain valid for all sample sizes.

DGP II. The second DGP is given by a model with fixed autoregressive matrix as in (31) and time varying volatility:

$$\begin{aligned} y_t &= By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Sigma_t), \quad y_0 = 0, \\ \Sigma_t &= H_t^{-1} \Psi_t H_t^{-1\prime} \end{aligned}$$

where H_t is a lower triangular matrix with ones on its main diagonal and bottom left element h_{21}, Ψ_t is a diagonal matrix, with diagonal elements, φ_{11} and φ_{22}, and

$$\begin{aligned} h_{21,t} &= \frac{\sum_{i=1}^{t} \xi_i}{\sqrt{t}}, \quad \xi_i \sim NID(0, 0.1^2), \\ \log \varphi_{kk,t} &= \frac{\sum_{i=1}^{t} \upsilon_i^k}{\sqrt{t}}, \quad \upsilon_i^k \sim NID(0, 0.1^2), \quad k = 1, 2. \end{aligned}$$

DGP II								
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
	QBLL_TVP_TVV	-0.0301	0.1532	0.9334	-0.0138	0.2186	0.9328
T=100	SS_TVP_TVV	-0.0168	0.1170	0.9508	-0.0051	0.2028	0.8389
	QBLL_FP_TVV	-0.0073	0.0801	0.9492	0.0232	0.2405	0.9564
	SS_FP_TVV	-0.0072	0.0986	0.8567	0.0013	0.1876	0.8211
	QBLL_TVP_TVV	-0.0137	0.0940	0.9416	-0.0057	0.1490	0.9401
T=500	SS_TVP_TVV	-0.0069	0.0710	0.9183	-0.0133	0.1347	0.8019
	QBLL_FP_TVV	-0.0023	0.0345	0.9492	0.0114	0.1546	0.9491
	SS_FP_TVV	-0.0021	0.0568	0.7483	-0.0118	0.1377	0.7963
	QBLL_TVP_TVV	-0.0110	0.0777	0.9414	-0.0093	0.1254	0.9326
T=1000	SS_TVP_TVV	-0.0063	0.0653	0.8626	-0.0187	0.1206	0.7900
	QBLL_FP_TVV	-0.0028	0.0260	0.9400	0.0048	0.1290	0.9424
	SS_FP_TVV	-0.0024	0.0512	0.7100	-0.0166	0.1254	0.7892

Table 2. Bias, RMSEs and coverage rates of models based on DGP II for 100, 500 and 1000 observations respectively.

The processes used for $h_{21,t}$, $\log \varphi_{11,t}$ and $\log \varphi_{22,t}$ resemble random walk processes, but are normalised by $\sqrt{t}$ in order to have finite variance, see Giraitis et al. (2014). In addition to the two models we fit to DGP I: QBLL_TVP_TVV and SS_TVP_TVV, we also include a QBLL model with fixed autoregressive parameters and time varying volatility, estimated using Algorithm 2 in Section 2.8 (QBLL_FP_TVV) and, similarly, a state space model with constant parameters and stochastic volatility (SS_FP_TVV). Table 2 displays the bias, RMSEs and coverage rate of the different models estimated with data generated with DGP II. It is clear from Table 2 that both QBLL_TVP_TVV and SS_TVP_TVV approaches perform well, with the state space model delivering: (i) smaller RMSEs when the sample size is small, for both the parameters and the volatility, but also, (ii) lower coverage rates (implying too narrow confidence intervals and the presence of type I errors) which seem to deteriorate with the sample size. Additionally, the two models featuring fixed autoregressive coefficients and drifting volatility: QBLL_FP_TVV and SS_FP_TVV models perform better than QBLL_TVP_TVV and SS_TVP_TVV respectively, particularly for the autoregressive parameters. This is to be expected, since both have a mechanism to build this additional information of parameter invariance in the estimation procedure.

DGP III. The third DGP is a model featuring both drifting parameters and volatility

$$y_t = B_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \Sigma_t), \quad y_0 = 0. \tag{32}$$

We use the same process for the volatility as in DGP II. In addition, the autoregressive parameters are generated as bounded random walks, subject to stability at each point in time:

$$B_{km,t} = \frac{\sum_{i=1}^{t} \zeta_i}{\sqrt{t}}, \qquad \zeta_i \sim NID(0, 0.1^2), \quad \text{for} \quad k, m = 1, 2.$$

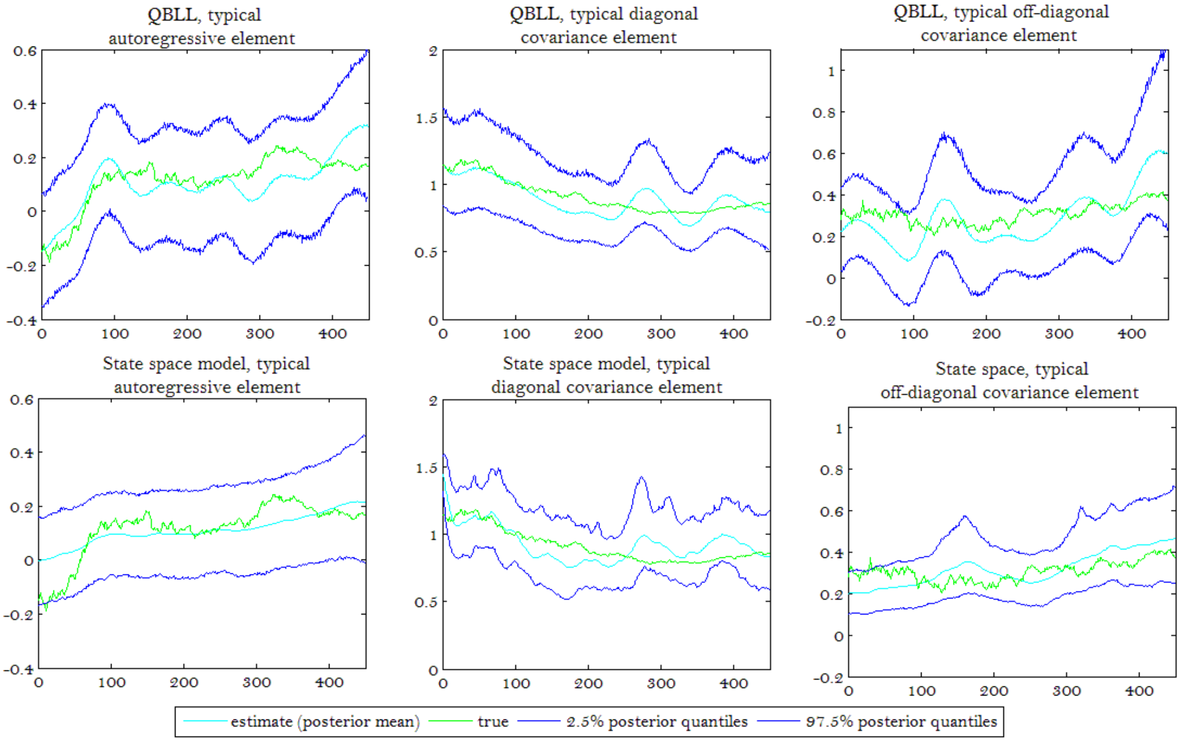
DGP III							
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
T=100	QBLL_TVP_TVV	-0.0046	0.1703	0.9473	-0.0275	0.2191	0.9350
	SS_TVP_TVV	-0.0058	0.1212	0.9135	0.0031	0.2121	0.8433
T=500	QBLL_TVP_TVV	-0.0001	0.1129	0.9449	-0.0106	0.1527	0.9423
	SS_TVP_TVV	0.0002	0.0662	0.8986	-0.0046	0.1398	0.8058
T=1000	QBLL_TVP_TVV	-0.0022	0.0911	0.9500	-0.0082	0.1247	0.9447
	SS_TVP_TVV	-0.0014	0.0515	0.8936	-0.0030	0.1229	0.8151

Table 3. Bias, RMSEs and coverage rates of models based on DGP III for 100, 500 and 1000 observations respectively.

Table 3 displays the bias, RMSE and coverage rate of the different models when the data are generated using DGP III. From Table 3, it clear that both the QBLL model and the state space model perform well; the state space model delivers smaller RMSEs particularly for the autoregressive parameters. The differences between the two models point estimates narrow when the sample size increases. Interestingly, the state space model's confidence intervals coverage rate deteriorates with the sample size, implying the presence of type I errors. We find that the confidence intervals implied by the state space model are sensitive to the tightness of the prior on the variance in the random walk state equations. We performed a robustness check and found that increasing the prior tightness delivers worse coverage rates, but loosing the prior, which is a way to 'let the data speak', considerably increases the difficulty of obtaining a stationary draw from the autoregressive matrix of the VAR and therefore the time required to estimate the model[8]. Figure 1 provides a graphical representation of the two estimation procedures for a typical realisation of the parameters and from Figure 1, the estimates of both methods look quite similar.

[8] In a robustness check, we removed the stability condition on the autoregressive matrix for all DGPs and loosened the prior on the state equation considerably. This delivered better coverage rates, coupled with a worse RMSE performance and a sizeable proportion of non-stable draws. These results are available upon request.

Figure 1: DGP III. Typical realisation of the time varying parameters and volatilities



DGP IV. The fourth DGP is given by a model (32), but here the autoregressive parameters and volatility follow the processes

$$
\begin{aligned}
B_{km,t} &= 0.25\sin(0.004\pi t) + 0.25\frac{\sum_{i=1}^{t}\zeta_i}{\sqrt{t}}, \qquad \zeta_i \sim NID(0, 0.3^2), \quad \text{for} \quad m, k = 1, 2, \\
h_{12,t} &= 0.5\sin(0.004\pi t) + 0.5\frac{\sum_{i=1}^{t}\xi_i}{\sqrt{t}}, \qquad \xi_i \sim NID(0, 0.3^2), \\
\log\varphi_{kk,t} &= 0.5\sin(0.004\pi t) + 0.5\frac{\sum_{i=1}^{t}\upsilon_i^k}{\sqrt{t}}, \qquad \upsilon_i^k \sim NID(0, 0.3^2), \quad k = 1, 2.
\end{aligned}
$$

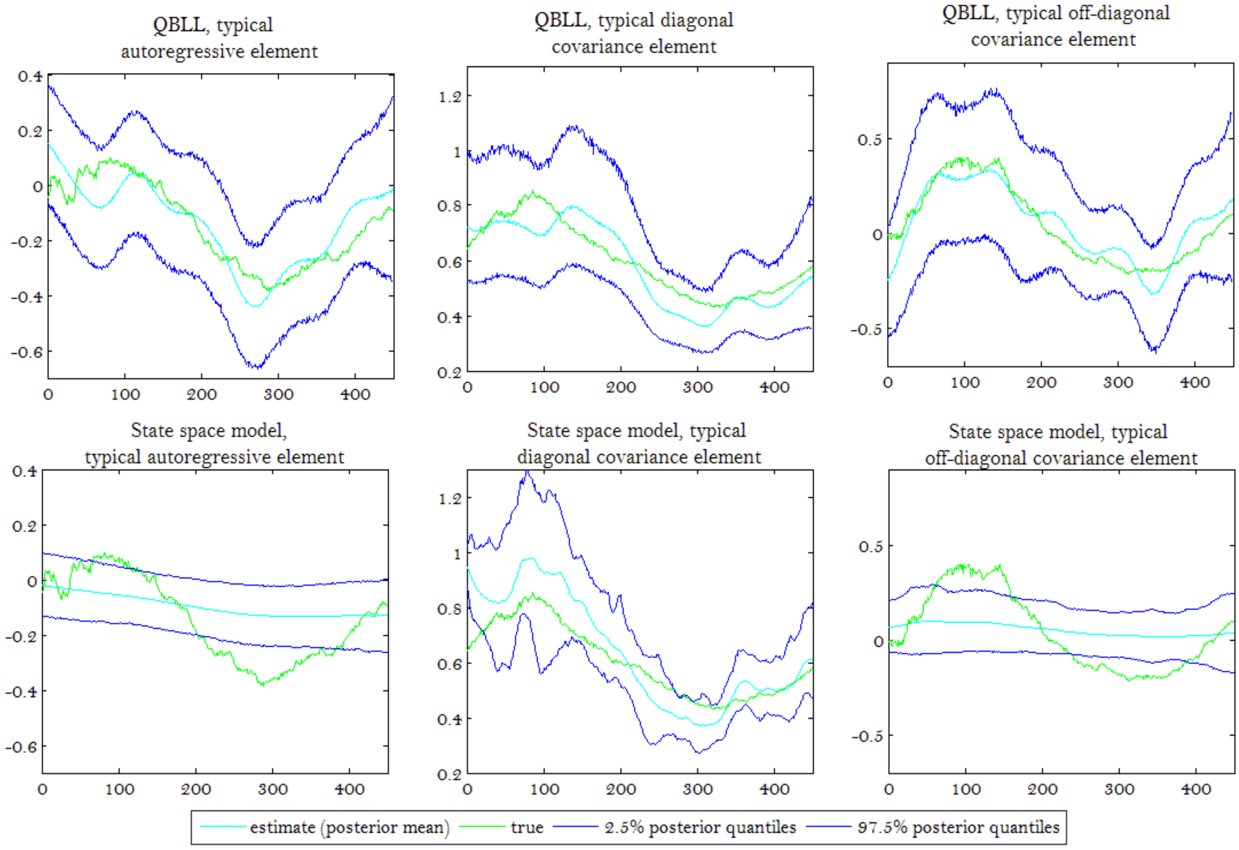
All parameters in DGP IV combine a deterministic time varying process following a sine wave and a stochastic part, generated by the bounded random walk process, used in DGPs II and III. For the estimation of the state space model, if the researcher knows ex-ante that the DGP contains a deterministic time trend, this information can be further added to the state equations. However, we continue using the random walk processes for the parameters, as these are widely used in the literature and the motivation behind our choice is to assess the consequences of using the 'wrong' state equation. Table 4 summarises the resulting bias, RMSE and coverage rate of the different models, including a fixed parameter BVAR model (F_BVAR) and two additional QBLL estimators, computed using different values of the bandwidth parameter H, when the sample is simulated using DGP IV.

		DGP IV					
		Autoregressive			Volatility		
		Bias	RMSE	Coverage	Bias	RMSE	Coverage
	QBLL_TVP_TVV, H=$T^{0.5}$	-0.0088	0.1853	0.9451	-0.0779	0.3657	0.9227
	QBLL_TVP_TVV, H=$T^{0.45}$	-0.0120	0.2034	0.9456	-0.0904	0.3982	0.9205
T=100	QBLL_TVP_TVV, H=$T^{0.55}$	-0.0060	0.1694	0.9435	-0.0679	0.3382	0.9238
	SS_TVP_TVV	-0.0094	0.1332	0.8922	-0.0372	0.3934	0.8040
	F_BVAR	0.0056	0.1303	0.8799	-0.0606	0.3673	0.7306
	QBLL_TVP_TVV, H=$T^{0.5}$	0.0057	0.1209	0.9470	-0.0204	0.2130	0.9313
	QBLL_TVP_TVV, H=$T^{0.45}$	0.0041	0.1384	0.9498	-0.0255	0.2402	0.9334
T=500	QBLL_TVP_TVV, H=$T^{0.55}$	0.0088	0.1074	0.9399	-0.0163	0.1949	0.9202
	SS_TVP_TVV	0.0302	0.1437	0.5176	-0.0349	0.3282	0.5560
	F_BVAR	0.0304	0.1366	0.3992	0.0384	0.1925	0.3789
	QBLL_TVP_TVV, H=$T^{0.5}$	0.0092	0.1004	0.9464	-0.0009	0.1833	0.9315
	QBLL_TVP_TVV, H=$T^{0.45}$	0.0035	0.1174	0.9485	-0.0100	0.2006	0.9418
T=1000	QBLL_TVP_TVV, H=$T^{0.55}$	0.0202	0.0930	0.9213	0.0104	0.1864	0.8666
	SS_TVP_TVV	0.0545	0.1812	0.2936	-0.0015	0.3613	0.4918
	F_BVAR	0.1127	0.2244	0.1697	0.1248	0.5559	0.0984

Table 4. Bias, RMSEs and coverage rates of models based on DGP IV for 100, 500 and 1000 observations respectively. The table also contains a comparison of the QBLL estimator with bandwidth parameters $T^{0.45}$, $T^{0.5}$ and $T^{0.55}$.

From Table 4 it is clear that both the fixed parameter model and the state space model provide invalid inference. More importantly, the problem does not vanish asymptotically, with RMSEs and coverage rates deteriorating with the sample size: the true autoregressive parameters are outside their 95% confidence intervals more than 70% of time implied by the state space model, and 80% of the time implied by the fixed parameter BVAR model for a sample size of 1000. This is evidence that parameter time variation is an issue that cannot be ignored since fixed parameter models in this case fail to deliver valid inference and are therefore an unsuitable choice. Moreover, we highlight once more the point that the state space model also fails to deliver valid inference, as it models the parameter process parametrically through the state equation and hence its statistical properties depend on the validity of the assumptions about the state equation. On the other hand, the nonparametric QBLL estimators remain valid in this case. Figure 2 further illustrates this last point by displaying typical realisations of the parameters over time: the state space model in this case does not account well for time variation and provides invalid confidence intervals. Table 4 also compares the QBLL estimator with bandwidth parameter $H = T^{0.5}$ with two alternatives: $H = T^{0.45}$ and $H = T^{0.55}$. From Table 4 we see that decreasing the bandwidth reduces the bias and increases the variance of the estimator. In general, there is a bias-variance trade-off and the bandwidth selection should depend on how slowly varying the underlying process of the parameters is, with larger bandwidths delivering smoother estimates. Giraitis et al. (2014) find that $H = T^{0.5}$

Figure 2: DGP IV. Typical realisation of the time varying parameters and volatilities



is asymptotically optimal (in the sense that it minimises the rate of the MSE of the estimator) but in DGP IV, we find that the estimator with slightly larger bandwidth $H = T^{0.55}$ delivers best RMSE performance, while the estimator with $H = T^{0.45}$ is best for coverage rates.

To summarise, in this Section we show that the class of QBLL estimators exhibit good finite sample properties, comparable and in some cases superior to those of state space models. The latter may suffer from type I error with probability increasing with the sample size, as well as non-decreasing RMSEs in case of misspecification of the state equation's parametric assumptions. We find that inference based on the quasi-posterior distributions introduced in Section 2.4 not only delivers valid confidence intervals in all cases examined but is also robust to different processes for the drifting parameters.

2.10 Applications to VAR models

2.10.1 Empirical application to U.S. monetary policy and inflation persistence

One question that has received a fair amount of attention and that divides the literature on monetary policy in the U.S. is whether the high inflation of the 1970's was an outcome of bad policy, suggesting the presence of time variation in the policy parameters of the model, or simply bad luck: a structural change in the size of shocks. In other words, was the subsequent period, referred to as the Great Moderation, a consequence of good policy under the leadership of Paul Volcker and Alan Greenspan or was it a good luck event - a decrease in the volatilities of the shocks hitting the economy. Supporters of the former view include Taylor (1993), DeLong (1997), Clarida, Gali and Gertler (2000) and Cogley and Sargent (2002, 2005). On the other hand, Sims (1980), Bernanke and Mihov (1998), Kim and Nelson (1999), McConnell and Perez Quiros (2000), Sims and Zha (2006) and Primiceri (2005) have insisted on explaining this phenomenon by the presence of changing volatility. A related question is whether inflation dynamics has changed after the Great Moderation period and, in particular, has inflation persistence decreased or remained constant over time as a consequence of the change in policy.

In this context, this Section contributes to the literature on changing monetary policy and inflation dynamics in the U.S. by applying the QBLL methodology proposed in this Chapter and revisiting classic results presented in Cogley and Sargent (2002, 2005), Primiceri (2005) and Cogley et al. (2010). Empirical contributions arise due to the novel approach of the Chapter which maintains an agnostic position on the parameter variation and, in turn, provides drifts in the parameters which are of nonparametric form and hence robust to state equation misspecifications. Additional contributions arise from the longer sample period, which spans from 1954Q3-2015Q3.

We estimate a structural time varying parameter BVAR(2) model with changing volatility on the macroeconomic series in Primiceri (2005)[9] employing the QBLL approach, introduced in Section 2.4. In addition, following Del Negro (2003)'s recommendation to include commodity prices due to their key role during the 1970's, the model is estimated with the addition of commodity prices[10] to the set of variables. The facilitation of the simple Minnesota prior on the autoregressive matrix at each time point allows to shrink directly the time varying elements and hence ensures stationary draws from the quasi-posterior. For the application, we use a Normal-inverted-Wishart conjugate

[9] The set of variables include annual GDP deflator inflation, civilian unemployment rate and the secondary market rate on the 3-month Treasury bills. Quarterly series for the unemployment rate and the nominal interest rate were computed as 3-month averages.

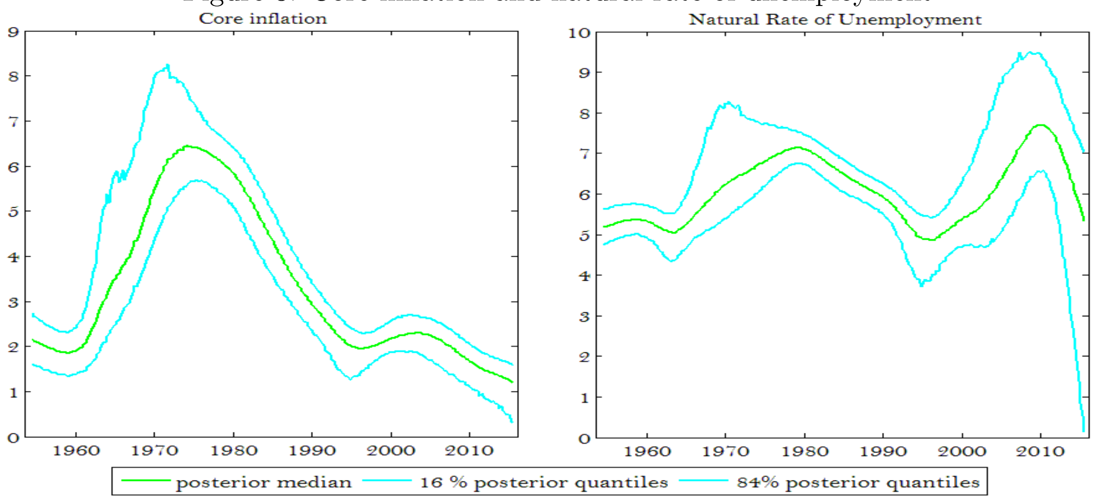
[10] Commodity prices are measured as annual growth of Moody's commodity price index.

prior with overall Minnesota shrinkage coefficient of 0.2 for the autoregressive parameters as in Bańbura et al. (2010) and prior for the Wishart parameters as in Kadiyala and Karlsson (1997)[11]. The model requires no stand on whether the Cholesky identifying restrictions are also time varying or not; this is an outcome of the estimation: the full contemporaneous covariance matrices are drawn from inverted-Wishart quasi-posterior and subsequently the Cholesky decomposition is employed at each point in time. Of course, for the structural analysis an assumption on the ordering of the variables is required in order to identify the monetary policy shock and we use the ordering in Primiceri (2005)[12]. First, we re-write the VAR model in equation (13) in companion form:

$$z_t = \mu_t + A_t z_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, \Omega_t). \tag{33}$$

Figure 3 presents the core inflation and natural rate of unemployment, which are computed as the

Figure 3: Core inflation and natural rate of unemployment



[11] A range of robustness checks were performed. First, our results are robust to different lag orders of the VAR model. In addition, the model was estimated with CPI inflation instead of GDP deflator; Fed Funds rate instead of the 3-month Treasury bill, GDP growth instead of unemployement and oil prices instead of commodity prices. The main results below do not change with these different specifications. Our results are also robust to different values of the overall shrinkage coefficient, λ. In particular, the model was estimated with $\lambda = 0.05$, 0.1, 0.5, 1 and 2 and the main results do not change.

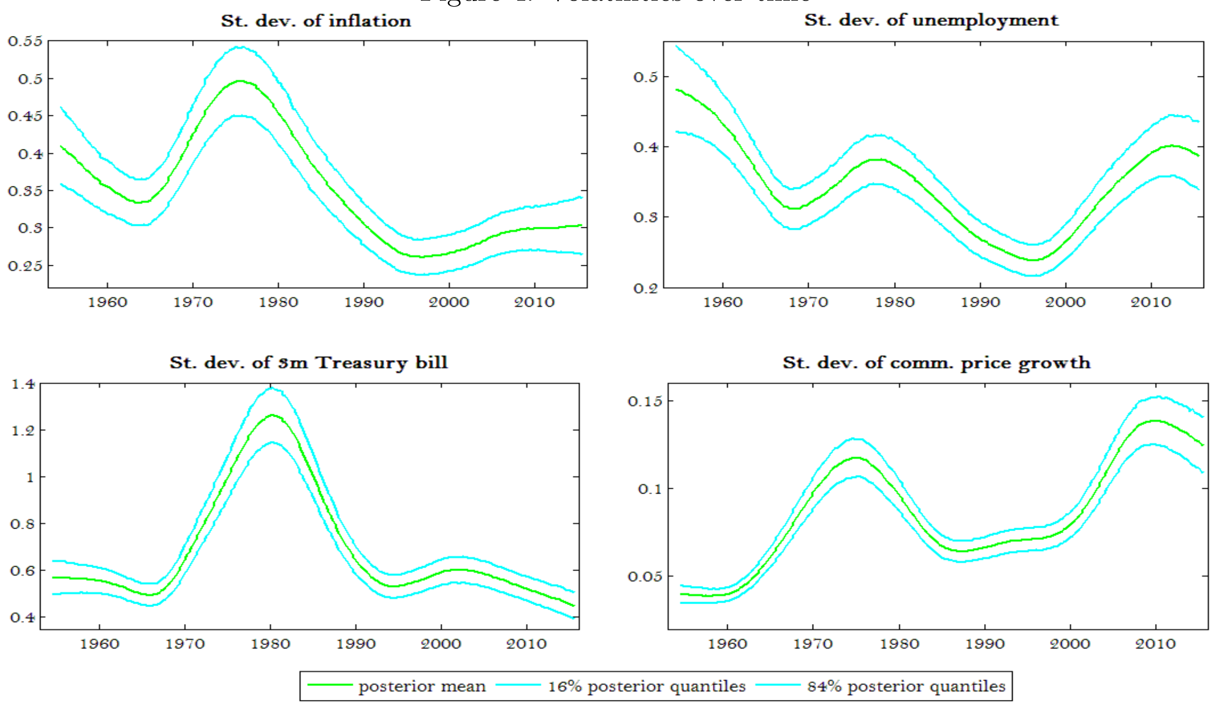
[12] In particular, the ordering is inflation, unemployment and the nominal interest rate last. In addition, commodity price is ordered first. The ordering of commodity price, inflation and unemployment is not an identifying assumption (alternative orderings in the non-policy block were estimated as robustness checks and our results do not change); however, the ordering of the nominal interest rate is crucial for identification and implies that monetary policy affects commodity price, inflation and unemployment with at least one quarter lag, see Christiano, Eichenbaum and Evans (1999) for details.

infinite horizon forecasts of inflation and unemployment implied by the model:

$$\tau_t = \lim_{h \to \infty} \mathbb{E}_t z_{t+1} \approx (I - A_t)^{-1} \mu_t,$$

where τ_t are the resulting stochastic trends of z_t. It is remarkable how similar the plots in Figure 3 are both quantitatively and in terms of shape to those presented in Cogley and Sargent (2002, 2005), even though their estimation method is considerably different. In particular, core inflation is around 2% in the 1960s, peaks to 6.5% in the 1970s and then falls down to pre-1970s levels after mid-1980s. The natural rate of unemployment also peaks in the 1970s but much less sharply, which is consistent with Cogley and Sargent (2002, 2005). More interestingly, since our sample contains the recent financial crisis, Figure 3 allows to assess the natural rate of unemployment during that period and it is evident that the increase is sharp and large in magnitude with the natural rate reaching 7.5 in 2009, when actual unemployment in the U.S. reached 10%. Figure 3 also suggests a recovery in the unemployment's natural rate after 2009. However, it is clear that the posterior confidence bands are much wider at the end of the sample implying larger uncertainty about its exact value after 2009. Figure 4 presents the volatilities of the four variables in the model over time[13]. These are

Figure 4: Volatilities over time



[13]The variation in the off diagonal elements of the covariance matrix can be found in the Appendix, Section 8.7.

consistent with the results presented in Primiceri (2005). Inflation volatility peaks in the mid-1970s and then falls dramatically during the Great Moderation period. Unemployment volatility is high in the late 1950s (note that most of the earlier papers do not use data during the period 1950-1960, which is typically used as pre-sample). Moreover, unemployment volatility peaks in the mid-1970s and is relatively low during the 1990s. We find that the 2008 financial crisis causes not only a large increase in the unconditional mean of unemployment but also in its variance, which remains high even after 2014. Finally, from Figure 4 we see that the 3-month Treasury rate's volatility exhibits a large increase in 1980 with aggressive monetary policy and tackling inflation under the leadership of Paul Volcker as chairman of the Fed and later falls to pre-1970s levels. Figure 5 compares the standard deviation of inflation from Figure 4 with its unconditional standard deviation, computed using the companion form of the model in (33) as

$$V(\pi_t) = e_\pi \left[\sum\nolimits_{j=0}^{\infty} A_t^j \Omega_t A_t^{j\prime} \right] e_\pi'$$

where e_π is a selection vector. It is worth comparing these results with Cogley et al. (2010) where they report the unconditional variance for different measures of inflation. While the standard deviation of inflation peaks only once in the late 1970s, from the right panel of Figure 5 it is clear that the unconditional standard deviation has a double peak, one around 1973 and another around 1979, coinciding with the two oil shocks. Both results are in line with evidence presented in Cogley et al. (2010).

Figure 5: Inflation Volatility

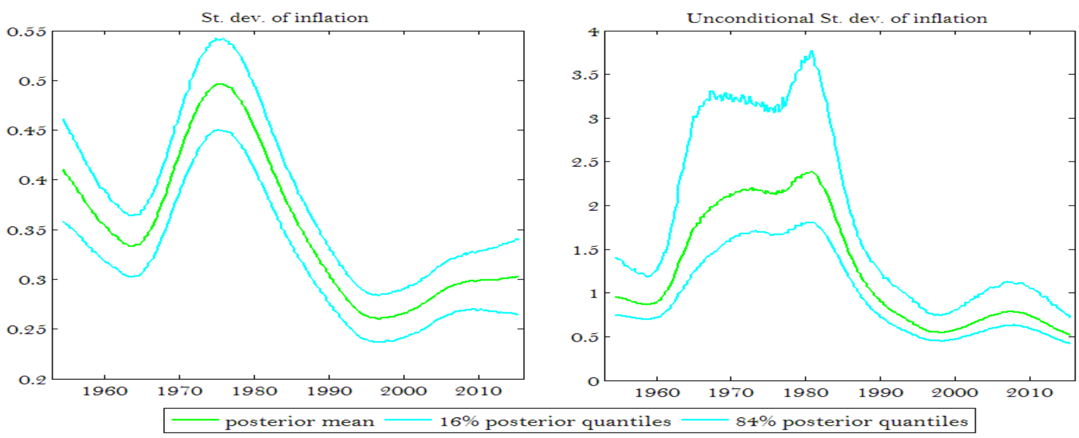
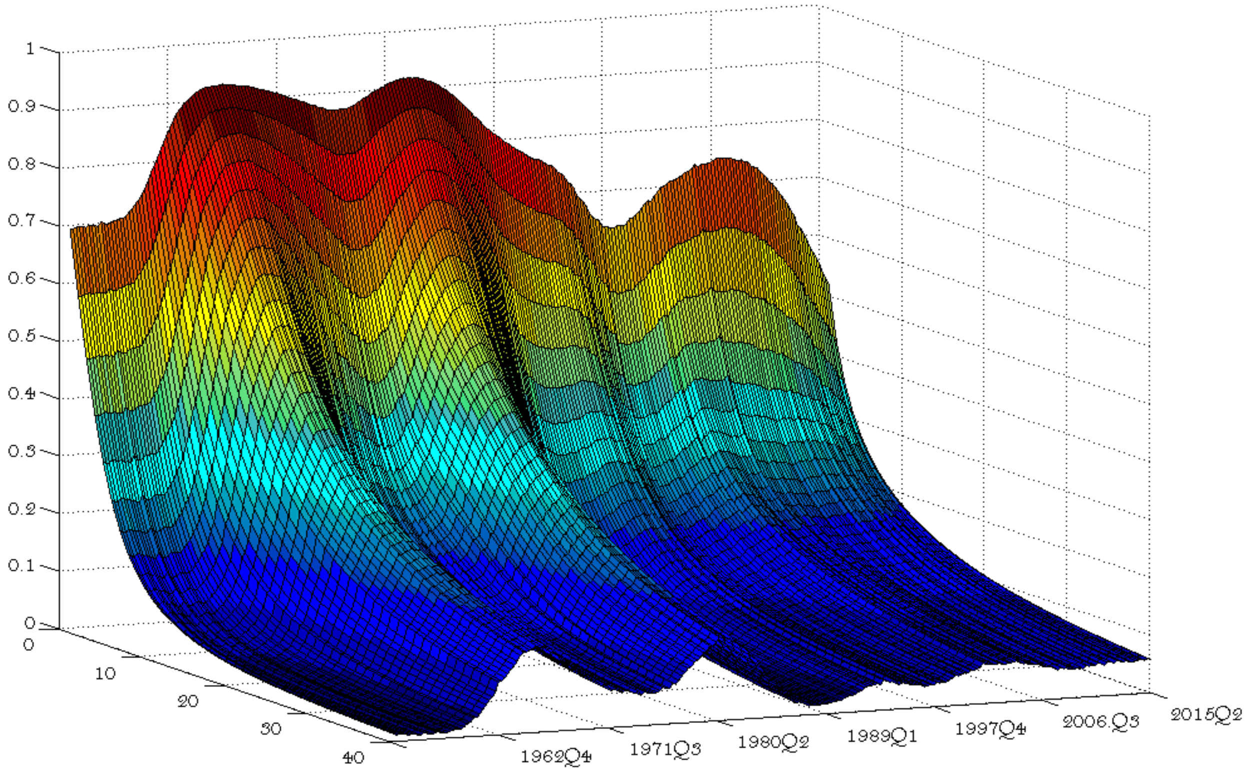


Figure 6: Inflation Persistence over time

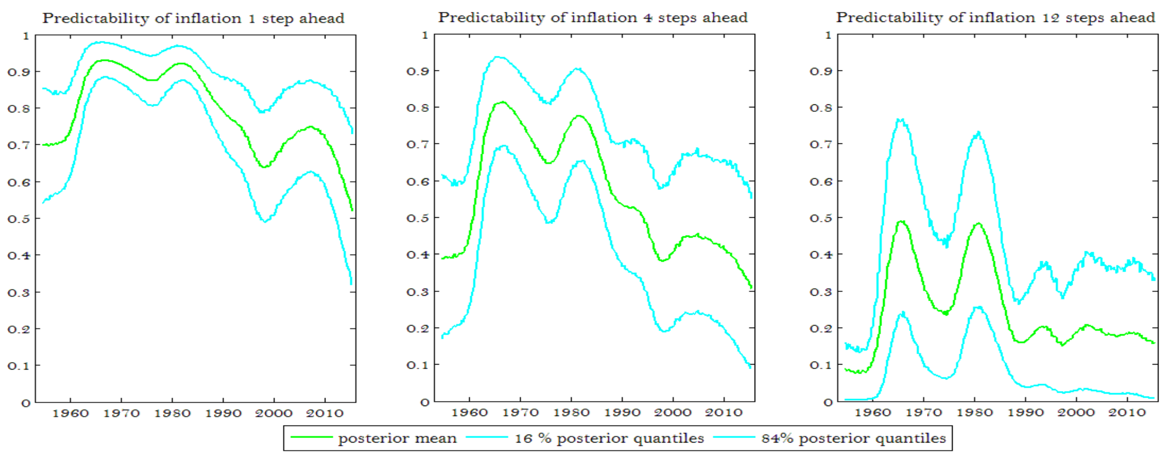


Next, we turn to the question of inflation persistence implied by the QBLL approach. Cogley and Sargent (2002) measure persistence as the normalised spectrum of inflation at zero. Cogley et al. (2010) introduce a new measure which they claim can be estimated more precisely. We follow their definition of inflation persistence h steps ahead as

$$R^2_{t,h} = 1 - \frac{e_\pi \left[\sum_{j=0}^{h-1} A_t^j \Omega_t A_t^{j\prime}\right] e_\pi'}{e_\pi \left[\sum_{j=0}^{\infty} A_t^j \Omega_t A_t^{j\prime}\right] e_\pi'}.$$

The measure $R^2_{t,h}$ represents the proportion of total variation explained by past shocks or equivalently one minus the proportion of total variation due to future shocks. It takes values between zero and one, with large values implying that past shocks die out slowly making inflation more persistent and hence predictable. Figure 6 presents the inflation persistence, measured by the posterior mean of $R^2_{t,h}$, computed at each point in time and for horizons ranging from one quarter to ten years ahead. In addition, Figure 7 displays the posterior mean of inflation persistence for selected horizons with 16% and 84% confidence bands. It is evident that inflation persistence dies out with lengthening the forecast horizon, which is expected. Furthermore, inflation is very persistent in the 1970s and early 1980 with two peaks roughly corresponding to the two oil shocks and persistence dies out slower over forecast horizons. The left panel of Figure 7 looks similar to what Cogley et al. (2010) report for inflation persistence for one quarter ahead forecasts. However, their TVP-VAR model delivers draws for $R^2_{t,1}$ which are very closely clustered at one around 1980 which they interpret as possible misspecification, while from the left panel of Figure 7, it is clear that the QBLL confidence bands for $R^2_{t,1}$ are bounded away from one. Another difference is that the confidence bands for $R^2_{t,h}$ in Figure 7, especially at medium horizons, are much more narrow than the ones presented in Cogley et al. (2010), delivering more precise estimates. Furthermore, due to the longer sample, we can assess inflation persistence during the 2008 financial crisis, when quarterly inflation reached negative values in 2009 and it is clear from both Figure 6 and 7 that inflation persistence increases during this period. However, unlike the period of 1970-1980s, in 2008 the $R^2_{t,h}$ measure dies out quickly over the forecast horizon and at 12 steps ahead there is no difference in inflation persistence during the crisis compared to previous periods.

Figure 7: Inflation Persistence



horizon	Inflation Persistence 1960-1980 p-value of increase	Inflation Persistence 1980-2006 p-value of decrease
1	0.916	0.899
4	0.901	0.886
12	0.928	0.860

Table 5. Posterior probabilities of change in inflation persistence for pair-wise periods. The table displays the probability of an increase of inflation persistence between 1960Q1 and 1980Q1 and of a decrease between 1980Q1 and 2006Q1 respectively.

To address the question of whether there is statistical evidence of higher persistence in 1980 compared to 1960 and statistically lower persistence in 2006 compared to 1980, Table 5 reports the posterior probabilities of an increase between 1960 and 1980 and those of a decrease between 1980 and 2006. The values in Table 5 suggest evidence of the presence of statistically significant change in inflation persistence for different horizons at levels close to 90% and this confirms the evidence presented in Cogley et al. (2010).

Next, we turn to the question of structural change in monetary policy. First, we can investigate the systematic component of monetary policy by computing the degree of monetary policy activism. Figure 8 displays the policy activism parameter, a_t, computed as: $a_t = \frac{b_t}{1-\rho_t}$, where b_t and ρ_t are the coefficients of one period lagged inflation and interest rate respectively in the interest rate equation of the VAR model. Periods in which a_t is below one can be interpreted as periods characterised by passive monetary policy or periods in which the Taylor principle is violated (see Woodford (2003) for details). From Figure 8, it is clear that the degree of policy activism implied by the QBLL approach is always above one, implying that we do not find evidence of passive monetary policy even in the periods of the oil crises. This result is at odds with evidence presented in Cogley and Sargent (2002, 2005) and since the QBLL approach is nonparametric with respect to the parameter processes, the time variation in a_t, and hence the validity of this result, do not depend on state equation assumptions. It is also evident from Figure 8 that during the first years of the appointment of Paul Volcker as a chairman of the Fed, the degree of policy activism is higher than other periods. This is consistent with a Taylor rule featuring changing parameters and supporting evidence is presented in Chapter 3 where we find in a linearised DSGE model with time varying parameters that there is an increase in the inflation targetting coefficient after the appointment of Paul Volcker. They also find that this increase coincides with a decrease in the output gap coefficient, suggesting that the Fed was giving priority to unemployment over the oil crises period.

Finally, we employ our time varying BVAR model for structural shock analysis in order to assess structural changes in the non-systematic component of monetary policy. We investigate whether there is evidence for changes in the transmission mechanism of the monetary policy shock to variables, resulting from the documented time variation in the parameters. Figure 9 displays the impulse response functions of inflation, unemployment and the nominal rate to a monetary policy shock. The top panel presents responses to a unit of the shock and hence measures only changes in the transmission of the shock over time without taking into account changes in the volatility. The bottom panel displays the responses to a one standard deviation of the shock and these incorporate the decrease in the volatility of the monetary policy shock over the second half

Figure 8: Monetary policy activism over time

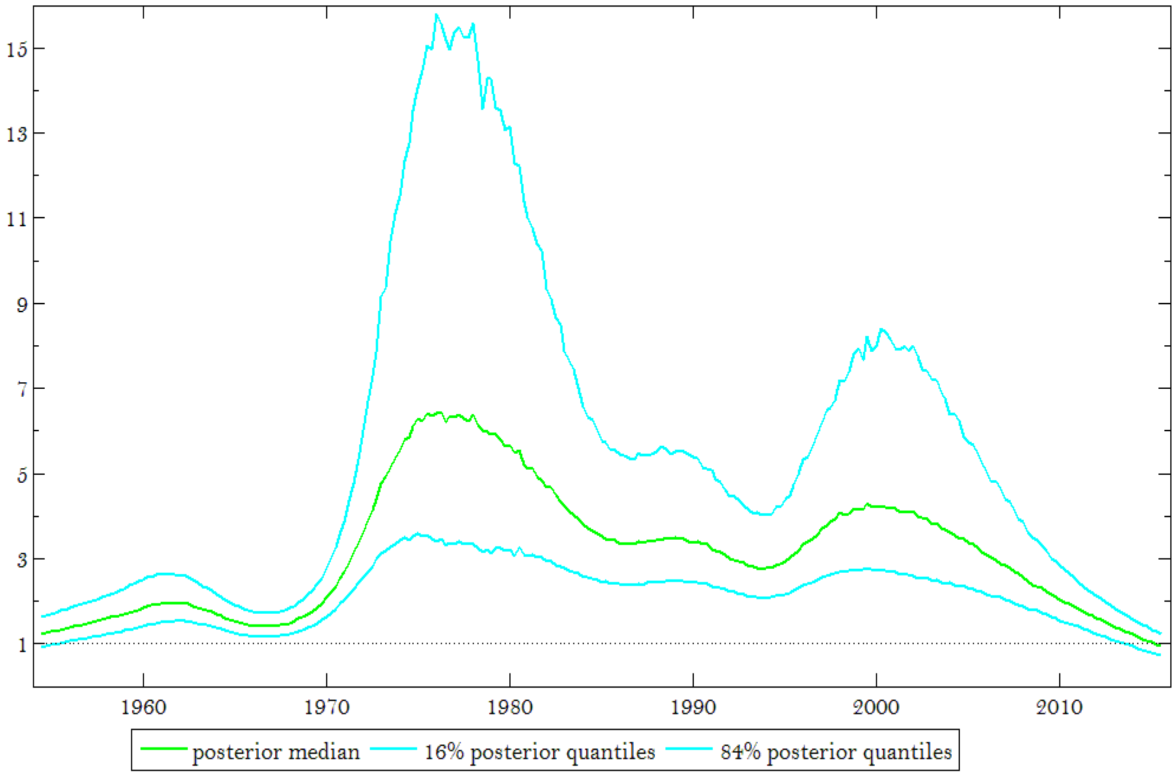
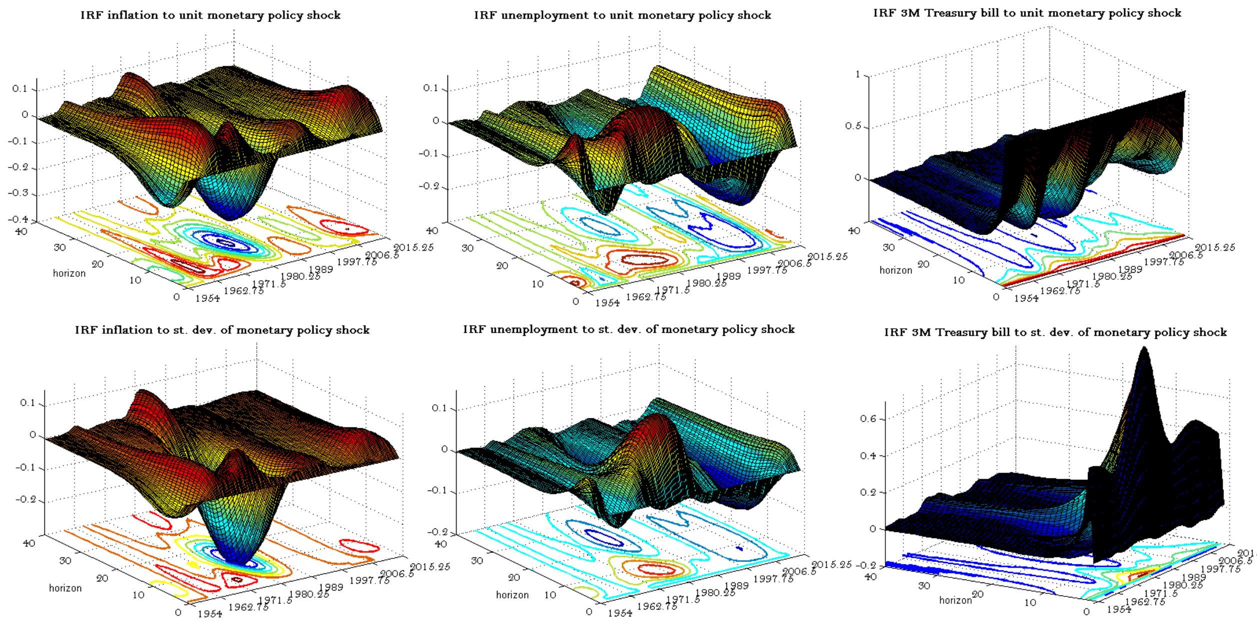
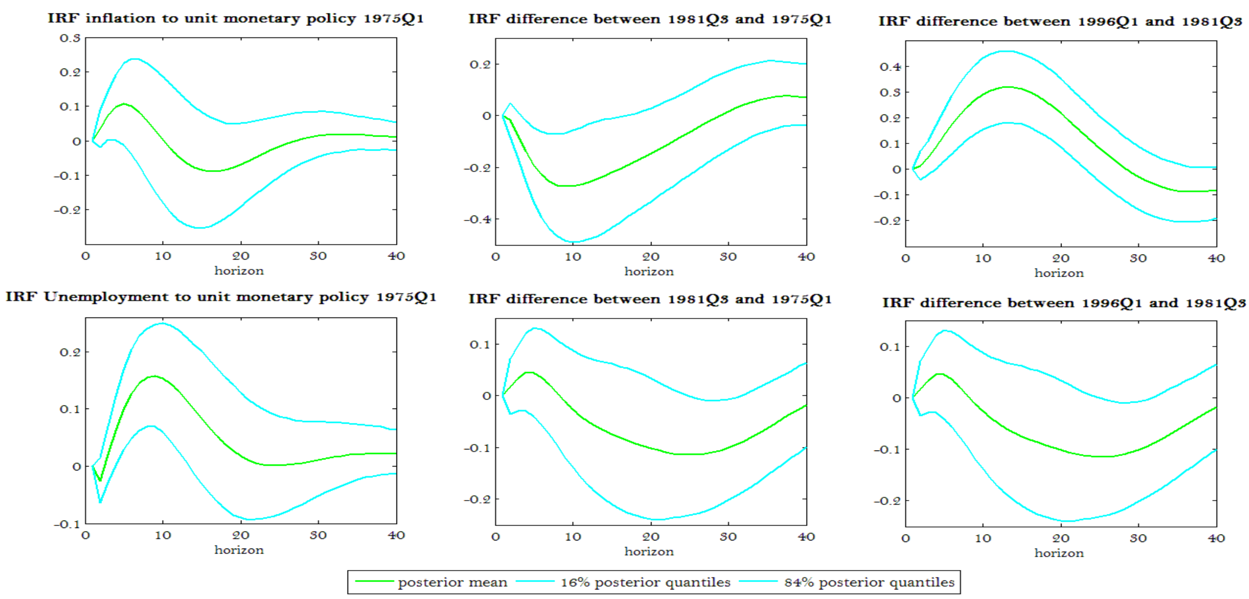


Figure 9: IRFs to monetary policy shock after including commodity prices



of the sample. From Figure 9, it is evident that inflation responds negatively to a unit shock of monetary policy apart from the periods around 1960 and 2008, i.e. there is a price puzzle during these periods: inflation increases with a positive monetary policy shock. Most of the puzzle goes away when considering the changing size of the shock over time and it is clear from the bottom left panel of Figure 9 that inflation responds substantially more to an unanticipated monetary policy shock during Paul Volcker's chairmanship than in any other period. Further evidence for

Figure 10: IRFs to unit monetary policy shock for selected periods



the increase in the inflation responsiveness to monetary policy shock can be found in the top panel of Figure 10, from where it is clear that the difference in inflation response between 1981Q3 and 1975Q1[14] is statistically significant. Interestingly, it appears from Figure 9 that inflation becomes less responsive to a monetary policy shock after the end of the second term of Paul Volcker in 1987 and from Figure 10 we also find statistically significant difference between inflation responses during Volcker's and Greenspan's years. Similarly, unemployment becomes less responsive to monetary policy shocks during the Great Moderation period. These results are consistent with evidence presented in Boivin and Giannoni (2006), who interpret the decreased responsiveness of inflation and output to a monetary policy shock after the 1980s which they also find in their application, as an outcome of the monetary authority becoming more effective and systematically more responsive

[14]The choice of dates for Figure 10 was motivated by Primiceri (2005), who uses the same periods for comparison. He argues that these periods are representative of the typical economic conditions of the chairmanships of Burns, Volcker and Greenspan, respectively.

in managing economic fluctuations after 1980. Finally, unemployment responds negatively to a positive policy shock during the recent crisis, which goes against standard monetary policy theory. This period is characterised by record high unemployment coupled with unconventional monetary policy, which cannot be captured by the nominal interest rate instrument frozen around the zero lower bound (ZLB). Hence, the model implied monetary policy shock is disabled during this period and cannot account for unemployment changes. This shortcoming can be addressed in several ways. One could to simply end the sample period at 2008 and this strategy has been employed by many in the literature. However, since the VAR model in (33) has a mechanism to account for structural change explicitly, the events in 2008 do not contaminate estimates in distant periods[15] and hence we choose to use the entire sample available. Alternatively, if we would like to say something about monetary policy in this period, then additional variables should be added to proxy for the unconventional monetary policy (for example, Baumeister and Benati (2013) identify an unconventional monetary policy shock as a shock to the spread of the 10 year Treasury yield over the nominal rate, while the nominal interest rate is unchanged at the ZLB). We leave this issue to future research and instead remain cautious in interpreting these results as anything other than misspecification due to omitted variables.

To summarise, we applied our novel QBLL approach to revisit much debated issues of changing macroeconomic dynamics, introduced in Cogley and Sargent (2002, 2005), Primiceri (2005) and Cogley et al. (2010). We found that not only does the volatility of the series change over time but also the autoregressive component of the VAR, implying that whereas there has been a change in policy, especially after Paul Volcker's appointment as chairman of the Federal Reserve, there has also been a 'luck' component implied by the considerable time variation uncovered in the volatility of the series. Our QBLL estimator delivers results broadly consistent with evidence presented in Cogley and Sargent (2002, 2005) and evidence of inflation persistence in Cogley et al. (2010), even though the estimation procedures differ considerably. In particular, unlike the parametric state space approach used in previous papers, the QBLL procedure is nonparametric with respect to the parameter processes and hence robust to various specifications. It is worth noting that the proposed methodology has the added advantage of being considerably simpler and computationally cheaper to implement and permits increasing the dimension of the VAR model to include additional relevant variables.

[15] In a robustness check, we estimated the model with sample ending in 2007Q4 and the estimation results for the pre-crisis periods do not change.

2.10.2 Forecasting exercise

In this Section we design a pseudo out-of-sample forecasting exercise in order to assess the forecasting record of the proposed QBLL approach. In particular, we compare the forecasts of various sizes BVAR models, with and without time varying parameters, using US data. The dataset[16] is from Stock and Watson (1996), spanning from 1950Q2 to 2014Q1. The dataset contains 87 quarterly macroeconomic series in first difference[17]. The forecast origins range from 1970Q2 to 2010Q2 and we compute forecasts for one up to eight quarters ahead. Accuracy of point forecasts is measured using the root mean squared forecast error (RMSFE) and forecast bias. Forecast density performance is evaluated using log predictive scores and probability integral transformation (PIT). The logscores and PITs are computed with the help of a nonparametric estimator to smooth the draws from the predictive density obtained for each forecast and horizon. We test whether a model is statistically more accurate than the benchmark against the two-sided alternative, with the Diebold and Mariano (1995)'s statistic computed with Newey-West estimator to obtain standard errors. The results of the Diebold-Mariano test are provided for the RMSFEs and logscores. For the bias, we test whether the models' bias is statistically different from zero. In addition, uniformity of the PITs is assessed using the test statistic of Berkowitz (2001), with a null hypothesis of uniform PITs.

We begin by estimating a set of small BVAR models which include 3 macroeconomic variables: GDP growth, inflation and the 3-month Treasury rate. The models are: a fixed parameter BVAR model (F_BVAR), a time varying parameter and volatility BVAR model estimated using the closed form QBLL expressions derived in Proposition 4 (QBLL_TVP_TVV), a drifting parameter homoscedastic BVAR model estimated with the Gibbs algorithm proposed in Section 2.8.1 (QBLL_TVP_FV) and an invariant parameter heteroscedastic BVAR estimated with the Gibbs algorithm developed in Section 2.8.2 (QBLL_FP_TVV). All BVAR models are of lag order one and use a Normal-inverted-Wishart conjugate prior with overall shrinkage for the autoregressive parameters as in Bańbura et al. (2010) and prior Wishart parameters as in Kadiyala and Karlsson (1997). In addition, the use of a small dataset permits comparison with alternative state space TVP-BVAR models. For the comparison, we estimate three different state space models: one with both autoregressive parameters and volatilities varying over time, one with invariant volatili-

[16] Detailed variable descriptions and data transformations can be found in Appendix 8.7.

[17] The choice of using a VAR model in first difference is motivated by results in Carriero, Galvão and Kapetanios (2015), who show in a forecast evaluation containing various models and countries, that their BVAR model in differences performs better on average than the one in levels.

ties and drifting parameters and one with invariant parameters and drifting volatilities[18], labelled SS_TVP_TVV, SS_TVP_FV and SS_FP_TVV respectively. Following the literature (Cogley and Sargent (2002), Primiceri (2005), Clark (2012), Mumtaz and Surico (2009)), the parameters, the elements of the lower triangular matrix and the log volatilities are all assumed to follow random walk processes.

	RMSFEs small models			Forecast bias small models		
horizon	GDP growth	Inflation	T-bill	GDP growth	Inflation	T-bill
			F_BVAR			
1	0.82	0.35	0.20	0.15*	-0.05*	0.01
2	0.84	0.46	0.21	0.18*	-0.05	0.01
4	0.85	0.56	0.20	0.18*	-0.03	0.01
8	0.84	0.64	0.20	0.19*	0.02	0.01
			SS_TVP_TVV			
1	1.01*	0.82*	1.00	0.16*	-0.03	0.01
2	1.00	0.80*	1.01	0.20*	-0.05	0.02
4	1.01	0.81*	1.00	0.21*	-0.08	0.02
8	1.01*	0.87*	1.05	0.23*	-0.08	0.02
			QBLL_TVP_TVV			
1	1.03	0.94	1.05*	0.03	0.03	0.00
2	1.03	0.85*	1.05*	0.03	0.05	0.00
4	1.05	0.79*	1.04	0.02	0.06	0.00
8	1.04	0.77	1.04*	0.04	0.08	0.00
			SS_FP_TVV			
1	1.02	0.82*	1.00	0.15*	-0.03	0.01
2	1.00	0.80*	1.02	0.20*	-0.04	0.02
4	1.01*	0.80*	0.99	0.21*	-0.05	0.02
8	1.01*	0.87*	1.00	0.21*	-0.04	0.02
			QBLL_FP_TVV			
1	0.98	0.87*	1.00	0.15*	-0.03	0.01
2	1.02	0.86*	0.99	0.19*	-0.05	0.01
4	1.02	0.89*	1.01	0.20*	-0.07	0.01
8	1.02*	0.97	1.00	0.20*	-0.06	0.01
			SS_TVP_FV			
1	1.02	0.83*	1.01	0.16*	0.00	0.02
2	1.00	0.81*	1.02*	0.20*	0.00	0.02
4	1.01	0.81*	0.99	0.20*	0.02	0.02
8	1.01	0.89*	1.00	0.20*	0.05	0.02
			QBLL_TVP_FV			
1	1.04*	0.95	1.07*	0.03	0.03	0.00
2	1.04	0.85*	1.08*	0.02	0.05	0.00
4	1.05	0.78*	1.03	0.02	0.06	0.00
8	1.04	0.77	1.06*	0.03	0.08	0.00

Table 6. RMSFEs and forecast bias. The figures in the left panel under F_BVAR are absolute RMSFEs, the figures under the remaining models are ratios of RMSFEs over the F_BVAR model. The figures in the right panel display forecast bias for each model. '*', '**' and '***' indicate rejection of the null of equal performance or zero forecast bias against the two-sided alternative at 10%, 5% and 1% significance level respectively.

The left panel of Table 6 presents the absolute performance of the small time invariant BVAR model (in RMSFEs) and the relative performance of alternative models over different horizons (numbers smaller than one imply superior performance of the particular model relative to F_BVAR). It

[18] We use versions of the algorithms outlined in Cogley and Sargent (2002, 2005) with 10,000 Gibbs draws from which the first 5,000 have been discarded. The priors and initial values are set using a presample of 30 observations.

is evident that the performance of all small models is very similar. In particular, they can rarely improve over the fixed parameter BVAR model for GDP growth and the 3-month Treasury rate, while all models deliver large and significant improvements in terms of RMSFEs for inflation. This is not surprising, given the evidence of considerable structural change in inflation dynamics presented in Section 2.10.1. The right panel of Table 6 presents the forecast bias of the different models over the sample (positive numbers imply positive bias) and one, two and three stars indicate rejection of the null of zero forecast bias at significance levels of 10%, 5% and 1% respectively.

From the right panel of Table 6, the fixed parameter BVAR displays a large significant positive bias for GDP growth. This is expected as the evaluation periods span from 1970Q2 up to 2010Q2 and arguably contain periods characterised by serious structural change. To give an illustration of the issue, in the beginning of the first oil crisis, the fixed parameter model delivers estimates by averaging the data from the pre-crisis insample period and even after the beginning of the crisis, it continues to overestimate output. On the other hand, time varying parameter models have a mechanism to capture the change in the relationships between variables and hence reduce bias. The QBLL approach also uses the insample period but progressively discounts distant data, giving more weight to the near past, and delivers unbiased forecasts. A more surprising result is that state space models featuring time variation in the parameters, and hence a mechanism to account for structural change, cannot improve systematic errors in the forecasts for GDP growth, implied by the statistically significant forecast bias. The gains of the QBLL_TVP_TVV over the state space models could be explained by recalling that unlike state space models, the QBLL approach does not force a parametric model on the parameter processes, which are allowed to vary freely and hence fit data better. Another interesting result is that while allowing the variances to change with the QBLL_FP_TVV model does not help with the bias, allowing for variation in the autoregressive parameters with the QBLL_TVP_FV model makes GDP growth forecasts unbiased. This is an indication that variation in the autoregressive component of the VAR is important for reducing forecast bias.

Table 7 accesses the quality of the density forecasts measured by logscores and PITs of the predictive density. The left panel of Table 7 displays absolute log predictive score for the F_BVAR model and differences in logscores for the alternative models, so numbers greater than zero imply superior performance over the F_BVAR. The right panel of Table 7 presents the p-values of the chi-squared Berkowitz (2001) test of uniformity of the PITs for all models (for example, numbers smaller than 0.05 indicate rejection of the null of uniform PITs at 5%). All models can deliver statistically significant improvements for inflation over the benchmark in terms of log predictive score. This is most likely a consequence of the superior point forecast implying that the forecast density is centered more precisely around the ex-post realised value. In addition, some models can also deliver superior forecast density performance for output growth and the interest rate. It is evident that the models with drifting volatility and fixed parameters SS_FP_TVV and QBLL_FP_TVV perform better than their counterparts, SS_TVP_FV and QBLL_TVP_FV. This can be further seen from the PIT p-values in the right panel of Table 7, reinforcing the view that for improving density forecasts, it is crucial to allow for changing volatility rather than variation in the autoregressive parameters.

	Log score small models			PITs small models		
horizon	GDP growth	Inflation	T-bill	GDP growth	Inflation	T-bill
			F_BVAR			
1	-1.24	-0.42	0.10	0.00	0.00	0.00
2	-1.28	-0.67	0.01	0.00	0.00	0.00
4	-1.28	-0.87	0.03	0.00	0.13	0.00
8	-1.29	-1.04	-0.02	0.00	0.64	0.00
			SS_TVP_TVV			
1	0.11***	0.33***	0.56***	0.02	0.73	0.23
2	0.08	0.42***	0.45***	0.00	0.02	0.54
4	0.03	0.44***	0.37**	0.00	0.00	0.40
8	0.00	0.37***	0.46*	0.00	0.01	0.02
			QBLL_TVP_TVV			
1	0.02	0.22***	0.23**	0.03	0.43	0.00
2	0.01	0.32***	0.18	0.01	0.04	0.00
4	-0.01	0.36***	0.14	0.00	0.00	0.00
8	-0.07	0.24	0.18	0.00	0.00	0.00
			SS_FP_TVV			
1	0.10***	0.33***	0.54***	0.03	0.95	0.45
2	0.10**	0.41***	0.44***	0.00	0.19	0.40
4	0.04	0.39***	0.36**	0.00	0.20	0.42
8	0.03	0.32***	0.47*	0.00	0.61	0.07
			QBLL_FP_TVV			
1	0.06	0.31***	0.33***	0.00	0.07	0.00
2	0.04	0.36***	0.27**	0.00	0.45	0.00
4	0.03	0.34***	0.09	0.00	0.13	0.00
8	0.02	0.20**	0.23	0.00	0.00	0.00
			SS_TVP_FV			
1	0.01	0.24***	0.05	0.01	0.79	0.00
2	0.03***	0.25***	0.08	0.00	0.24	0.00
4	0.01	0.21***	0.08	0.00	0.39	0.00
8	0.02	0.03	0.06	0.00	0.00	0.00
			QBLL_TVP_FV			
1	-0.04**	0.10***	-0.08	0.00	0.00	0.00
2	-0.01	0.24***	-0.08	0.01	0.01	0.00
4	-0.02	0.34***	0.00	0.01	0.18	0.00
8	0.00	0.39**	0.05	0.00	0.03	0.00

Table 7. Log predictive scores and PITs. The figures in the left panel under F_BVAR are absolute log scores, the figures under the remaining models are differences of log scores from the F_BVAR model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test. The figures in the right panel display p-values of the chi-squared Berkowitz (2001)'s test of uniformity for all models, values in bold indicate models that cannot reject the null of uniformity at 95%.

horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
	RMSFEs medium and large models									
	F_BVAR									
1	0.77	0.34	0.40	0.03	0.73	0.07	1.25	0.20	0.85	0.50
2	0.82	0.41	0.53	0.04	0.71	0.08	1.60	0.20	0.86	0.65
4	0.84	0.50	0.59	0.04	0.68	0.09	1.57	0.20	0.88	0.72
8	0.84	0.62	0.70	0.04	0.69	0.09	1.59	0.20	0.88	0.85
	SS_TVP_TVV									
1	1.08*	0.85*	-	-	-	-	-	1.00	-	-
2	1.03	0.89*	-	-	-	-	-	1.02	-	-
4	1.01	0.91*	-	-	-	-	-	1.00	-	-
8	1.02	0.90*	-	-	-	-	-	1.05	-	-
	QBLL_TVP_TVV									
1	1.04	0.82*	0.87*	1.10*	0.86*	1.05	1.01	1.07	1.05	0.93
2	1.02	0.84*	0.88*	1.01	0.96	1.07	1.01	1.11*	0.99	0.95
4	1.05	0.79*	0.85*	1.08*	1.06	1.11*	1.10*	1.03	0.99	0.93
8	1.05	0.78	0.83	1.07*	1.08*	1.14*	1.08*	1.05*	1.01	0.90
	QBLL_FP_TVV									
1	1.00	0.86*	0.97	1.14*	0.93	1.10	0.96	1.06	0.99	0.93*
2	1.01	0.84*	0.92*	0.99	0.93	1.03	0.99	0.97	0.99	0.92*
4	0.96*	0.81*	0.88*	0.98	0.97	1.02	0.99	1.00	1.01	0.91*
8	0.99	0.93*	0.96	1.01	1.01	1.00	0.97*	0.99	1.02	0.98
	QBLL_TVP_FV									
1	1.11*	0.99	0.94	1.26*	0.85*	1.20*	1.16*	1.26*	1.01	1.01
2	1.05	0.93	0.88*	1.13	0.95	1.13*	1.05	1.18*	0.96	0.98
4	1.06	0.87	0.91	1.03	1.03	1.10*	1.04	1.14	0.99	0.95
8	1.07*	0.80	0.85	1.27*	1.01	1.32*	1.07*	1.27*	0.98	0.90
	F_BVAR_L									
1	0.96	0.94	0.99	0.99	0.90*	0.91	0.96	1.00	1.03	1.00
2	0.92*	0.93*	0.98	0.91*	0.93*	0.95	0.90	0.97	0.99	1.00
4	1.01	0.96	0.98	1.01	0.98	1.03	1.02	1.00	1.00	1.00
8	0.99	0.95*	0.96*	0.99*	1.00	1.01	0.99	1.01*	1.00	0.97
	QBLL_TVP_TVV_L									
1	1.01	0.83*	0.92	1.14*	0.85*	0.96	0.99	1.07	1.09*	0.99
2	1.00	0.81*	0.89*	1.02	0.93	1.02	0.99	1.09*	0.99	0.96
4	1.08*	0.81*	0.87*	1.11*	1.07	1.14*	1.14*	1.04	1.00	0.95
8	1.08*	0.79	0.84	1.09*	1.09*	1.21*	1.13*	1.06*	1.02	0.90

Table 8. RMSFEs. The figures under F_BVAR are absolute RMSFEs, the figures under the remaining models are ratios of RMSFEs over the F_BVAR model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

Next, we investigate whether including additional variables to the model can improve the forecast performance. This choice is motivated by the forecasting literature (Bańbura et al. (2010), Koop (2011), Carriero, Clark and Marcellino (2015a)), which suggests good forecasting record of large dimensional BVAR models. In particular, we estimate two models: (i) a fixed parameter model (F_BVAR), and (ii) a time varying parameter model with changing volatility (QBLL_TVP_TVV), on a medium size dataset (consisting of 17 macroeconomic variables described in Appendix 7.1.10). Using the medium dataset, we also estimate two mixture of time varying and time invariant parameter models using Algorithms 1 and 2 developed in Sections 2.8.1 and 2.8.2 respectively: (i) QBLL_FP_TVV is a BVAR model with time invariant parameters and time varying volatility, and (ii) QBLL_TVP_FV is a BVAR model with time varying autoregressive coefficients and homoscedastic covariance matrix. Finally, we estimate two large BVAR models[19] using the entire dataset (consisting of 87 variables): (i) a model with time invariant parameters labelled F_BVAR_L, and (ii) a model, estimated with the QBLL approach, featuring time varying parameters and time varying volatility (QBLL_TVP_TVV_L).

	Forecast bias medium and large models									
horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
	F_BVAR									
1	0.14*	-0.01	-0.05	0.00	0.07	0.00	0.24*	0.01	0.17*	-0.08*
2	0.16*	-0.02	-0.06	0.00	0.11*	0.00	0.27*	0.01	0.21*	-0.11*
4	0.17*	-0.03	-0.07	0.00	0.13*	0.00	0.28	0.01	0.22*	-0.13
8	0.19*	0.00	-0.05	0.00	0.16*	0.00	0.33*	0.01	0.22*	-0.11
	SS_TVP_TVV									
1	0.16*	-0.03	-	-	-	-	-	0.01	-	-
2	0.20*	-0.05	-	-	-	-	-	0.02	-	-
4	0.21*	-0.08	-	-	-	-	-	0.02	-	-
8	0.23*	-0.08	-	-	-	-	-	0.02	-	-
	QBLL_TVP_TVV									
1	0.05	0.02	0.03	0.00	0.06	-0.01	0.07	0.00	0.03	0.03
2	0.02	0.04	0.04	0.00	0.05	0.00	0.03	0.00	0.01	0.04
4	0.01	0.05	0.04	0.00	0.04	0.00	-0.04	0.00	0.02	0.05
8	0.03	0.07	0.06	0.00	0.05	0.00	-0.01	0.00	0.03	0.06
	QBLL_FP_TVV									
1	0.10*	0.00	-0.06*	0.00	0.12*	0.01	0.19*	0.01	0.17*	-0.06*
2	0.10	-0.01	-0.07*	0.00	0.13*	0.00	0.17	0.00	0.22*	-0.10*
4	0.10	-0.04	-0.09	0.00	0.15*	0.01	0.16	0.00	0.23*	-0.12
8	0.12	-0.05	-0.11	0.00	0.18*	0.01	0.20	0.00	0.24*	-0.15
	QBLL_TVP_FV									
1	0.03	0.01	0.02	0.00	0.07	0.00	0.02	0.01	0.04	0.02
2	0.04	0.03	0.04	0.00	0.05	0.00	0.08	0.01	0.02	0.05
4	0.03	0.05	0.05	0.00	0.04	0.00	-0.01	0.00	0.03	0.06
8	0.03	0.07	0.07	0.00	0.07	0.01	0.01	0.00	0.03	0.06
	F_BVAR_L									
1	0.01	-0.06*	-0.07*	0.00	0.02	0.01*	0.01	0.00	0.08	-0.11*
2	0.07	-0.08*	-0.12*	0.00	0.08	0.01	0.10	-0.01	0.15*	-0.18*
4	0.16*	-0.11	-0.15*	0.00	0.14*	0.00	0.26	0.00	0.21*	-0.22*
8	0.20*	-0.08	-0.12	0.00	0.18*	0.00	0.36*	0.01	0.22*	-0.19
	QBLL_TVP_TVV_L									
1	0.02	0.02	0.02	0.00	0.04	0.00	0.02	0.00	0.02	0.01
2	0.01	0.04	0.03	0.00	0.03	0.00	-0.01	0.00	0.02	0.03
4	-0.01	0.05	0.04	0.00	0.02	0.01	-0.09	0.00	0.03	0.05
8	0.00	0.07	0.05	0.00	0.04	0.01	-0.07	-0.01	0.03	0.06

Table 9. Forecast bias. The figures display forecast bias for each model. '*', '**' and '***' indicate rejection of the null of zero forecast bias against the two-sided alternative at 10%, 5% and 1% significance level respectively.

Table 8 presents the absolute performance of the medium time invariant BVAR model (in RMSFEs) and the relative performance of alternative models over different horizons. From Table 8, it is clear that even a time invariant model with additional variables can outperform the small time varying model, SS_TVP_TVV, for GDP growth. In addition, the various time varying medium

[19] As explained in Bańbura et al. (2010), the overall shrinkage, λ, should be invertly proportional to the size of the model. The shrinkage parameter for the small models is 0.3, for the medium models 0.2, and for the large models 0.11. A grid for other values of λ were considered for each model and size and whereas the absolute forecast performance of the models depends on λ, the relative models' performance presented in this section is robust to different values of λ.

and large models can outperform the SS_TVP_TVV for inflation implied by the larger improvements over the fixed parameter model. The best performing model for inflation is the medium size QBLL_TVP_TVV model. In addition, the time varying specifications can improve RMSFE performance of the F_BVAR for consumption growth and PCE growth.

	Log score medium and large models									
horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
	F_BVAR									
1	-1.19	-0.35	-0.51	2.03	-1.08	1.30	-1.67	0.08	-1.28	-0.70
2	-1.25	-0.56	-0.79	1.77	-1.13	1.08	-1.93	-0.02	-1.37	-0.96
4	-1.28	-0.79	-0.94	1.76	-1.13	0.99	-1.96	-0.03	-1.37	-1.16
8	-1.29	-0.99	-1.23	1.75	-1.14	0.98	-1.97	-0.02	-1.37	-1.35
	SS_TVP_TVV									
1	0.05	0.26***	-	-	-	-	-	0.59***	-	-
2	0.04	0.31***	-	-	-	-	-	0.48***	-	-
4	0.03	0.35***	-	-	-	-	-	0.43**	-	-
8	0.00	0.33***	-	-	-	-	-	0.45*	-	-
	QBLL_TVP_TVV									
1	0.07	0.29***	0.19**	-0.03	0.19***	-0.01	0.12**	0.29**	0.08	0.13
2	0.02	0.35***	0.15**	0.00	0.15***	-0.11	0.04	0.24*	0.13	0.07
4	0.03	0.39***	0.09	-0.04	0.05	-0.22	-0.11	0.22	0.11	0.11
8	0.00	0.33**	0.07	-0.11	-0.04	-0.34	-0.08	0.22	0.10	-0.01
	QBLL_FP_TVV									
1	0.06	0.20***	-0.01	-0.08	0.09**	-0.08	0.17***	0.32**	0.06	0.04
2	0.07	0.27***	0.08**	0.01	0.08	-0.05	-0.04	0.34**	0.13	0.10**
4	0.12*	0.32***	0.18***	0.01	0.11*	-0.05	-0.01	0.22*	0.12	0.18**
8	0.07	0.12	0.03	-0.03	0.04	-0.03	0.06	0.22	0.09	-0.03
	QBLL_TVP_FV									
1	-0.08	0.13***	0.07	-0.22***	0.07	-0.19***	-0.24***	-0.25**	-0.05	0.01
2	-0.05*	0.18***	0.08	-0.18***	0.06	-0.18***	-0.08*	0.04	0.04	-0.03
4	-0.03	0.25***	0.09	-0.23***	0.03	-0.21***	-0.03	0.09	0.04	0.13
8	-0.06**	0.32*	0.40*	-0.34***	-0.01	-0.33***	-0.02	-0.08	0.09	0.27*
	F_BVAR_L									
1	0.07	0.07***	-0.02	0.01	0.07**	0.07**	0.04**	0.07	-0.04	0.01
2	0.08***	0.08***	0.03	0.09***	0.05**	0.07*	0.10***	0.08	0.05	0.00
4	0.00	0.07***	0.05*	0.00	0.01	-0.01	0.00	0.08	0.04	0.02
8	0.01	0.05**	0.07**	0.03*	-0.01	-0.02	0.00	-0.06	0.01	0.03
	QBLL_TVP_TVV_L									
1	0.09	0.30***	0.07	-0.06	0.20***	0.09*	0.16***	0.38***	0.04	-0.04
2	0.08*	0.37***	0.16**	0.05	0.17***	0.06	0.13*	0.31**	0.12	0.07
4	-0.02	0.39***	0.09	-0.04	0.05	-0.16	-0.07	0.22	0.11	0.12
8	-0.03	0.35**	0.09	-0.12	-0.04	-0.37	-0.11	0.16	0.10	0.06

Table 10. Log predictive scores. The figures under F_BVAR are absolute log scores, the figures under the remaining models are differences of log scores from the F_BVAR model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

Table 9 presents the forecast bias of the different models. It is evident that, while statistically significant forecast bias is present in medium and large time invariant models, the forecasts of the medium and large time varying specifications are virtually unbiased for all variables and all horizons. More interestingly, when we consider the medium model with drifting volatility and time invariant autoregressive component, we also find forecast bias in some variables. On the other hand, allowing for drifting autoregressive parameters when the volatility is held constant over time removes the bias. This confirms the results from the small models: variation in the autoregressive component of the VAR can significantly reduce systematic errors in the forecasts. Table 10 accesses the quality of the density forecasts of the medium and large models. It displays absolute log predictive score for the F_BVAR model and differences in logscores for the alternative models. We see from Table 10 that the various time varying specifications can deliver forecast density improvements for most variables. In particular, the best performing model is the large time varying parameter and volatility model, QBLL_TVP_TVV_L, implying that by including additional variables we can obtain superior estimates of the uncertainty around the point forecast.

	PITs medium and large models									
horizon	GDP Growth	Inflation (deflator)	PCE Growth	Inv. Growth	Cons. Growth	Unemp. Rate	Ind. Production	T-bill	Real Wage Growth	Inflation (CPI)
	F_BVAR									
1	0.04	0.00	0.09	0.00	0.00	0.11	0.00	0.00	0.00	0.00
2	0.00	0.00	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.13
8	0.00	0.75	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.05
	SS_TVP_TVV									
1	0.02	0.73	-	-	-	-	-	0.23	-	-
2	0.00	0.02	-	-	-	-	-	0.54	-	-
4	0.00	0.00	-	-	-	-	-	0.40	-	-
8	0.00	0.01	-	-	-	-	-	0.02	-	-
	QBLL_TVP_TVV									
1	0.61	0.32	0.52	0.93	0.13	0.35	0.56	0.10	0.07	0.59
2	0.17	0.25	0.00	0.14	0.24	0.00	0.00	0.34	0.09	0.00
4	0.42	0.06	0.00	0.38	0.29	0.00	0.00	0.00	0.15	0.00
8	0.14	0.00	0.00	0.00	0.01	0.00	0.00	0.02	0.09	0.00
	QBLL_FP_TVV									
1	0.10	0.04	0.00	0.01	0.01	0.00	0.06	0.00	0.00	0.00
2	0.15	0.48	0.01	0.27	0.02	0.00	0.00	0.68	0.00	0.00
4	0.12	0.03	0.20	0.08	0.00	0.00	0.00	0.00	0.00	0.03
8	0.08	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.00	0.00
	QBLL_TVP_FV									
1	0.00	0.01	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.03
2	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00
4	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.71	0.25
	F_BVAR_L									
1	0.05	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
8	0.00	0.04	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00
	QBLL_TVP_TVV_L									
1	0.81	0.16	0.03	0.82	0.12	0.76	0.22	0.00	0.10	0.00
2	0.72	0.13	0.00	0.94	0.12	0.38	0.62	0.44	0.19	0.00
4	0.13	0.14	0.00	0.66	0.60	0.00	0.00	0.05	0.38	0.00
8	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.02	0.29	0.00

Table 11. PITs. The figures display p-values of the chi-squared Berkowitz (2001)'s test of uniformity for all models, values greater than 0.05 indicate that we cannot reject the null of uniformity at 95%.

Finally, Table 11 presents the p-values of the chi-squared Berkowitz (2001) test of uniformity for all medium and large models. From Table 11, it is clear that for time invariant models, both medium and large, we can reject the null of uniform PITs for most variables and horizons. On the other hand, for the medium time varying model, QBLL_TVP_TVV, we cannot reject the null of uniformity for all variables one step ahead. The large time varying model, QBLL_TVP_TVV_L, also performs exceptionally well. Another important result is that the model featuring drifting volatility and invariant autoregressive parameters performs better in terms of PITs than the model with drifting parameters and invariant volatility, confirming our previous result that for density forecast performance, drifting volatility is more important than drifting autoregressive parameters.

To summarise, a forecasting exercise was conducted, where we estimated various sizes fixed and time varying parameter BVAR models. Several conclusions emerged: first, the small QBLL models can deliver model forecast performance similar to that of state space models. Second, the QBLL approach allows to increase the number of variables in the system: as we demonstrated increasing the VAR dimension while allowing for time variation in the parameters can improve both point and density forecasts over (i) small models, and (ii) invariant models of the same dimensions. Third, the QBLL approach performs exceptionally well in eliminating forecast bias as well as in delivering uniform PITS and this is true for most variables and horizons. Finally, by assessing the forecast performance of BVAR models with mixtures of time varying and time invariant parameters, we reached an important conclusion: allowing for drifting volatility in a BVAR model improves density forecasts compared to a model with invariant volatility; on the other hand, time variation in the autoregressive component of the VAR model can significantly reduce forecast bias. Since both point and density performance is crucial for forecasting, we conclude that variation in both the parameters and variances of BVAR models should be considered.

2.11 Summary

This Chapter establishes a novel quasi-Bayesian local likelihood (QBLL) approach for econometric inference in models with time varying parameters. The Bayesian framework is based on augmenting the local likelihood of Giraitis et al. (2016) with a prior distribution; this augmentation principle delivers asymptotically valid quasi-posterior distributions which admit closed form expressions in the special case of a linear Gaussian univariate regression model and multivariate VAR model. The approach is of sufficient generality and flexibility to give rise to Gibbs algorithms that are able to sample from a BVAR model with a mixture of time varying and time invariant parameters.

The Monte Carlo exercise demonstrates that the class of QBLL estimators based on: (i) the closed form joint distribution derived in Section 2.4; (ii) the proposed MCMC algorithms in Section 2.8, both exhibit good finite sample properties, and inference based on their quasi-posterior densities delivers valid confidence intervals. Importantly, QBLL inference is robust to different processes for the drifting parameters, as its validity does not depend on parametric restrictions typically imposed by state space models.

The novel QBLL approach is employed to empirically address the issue of changing macroeconomic dynamics in the U.S. and confirms previous results of Cogley and Sargent (2002, 2005) and Primiceri (2005) on the presence of significant structural change in core inflation and the natural rate of unemployment as well as of substantial drifts in the volatility of the series. In the light of these results, we conclude that ignoring the time variation (by instead estimating a time invariant model) will result in invalid inference on the model's parameters. We also find that inflation has become significantly less persistent after the beginning of the Great Moderation, which is in line with more recent results in Cogley et al. (2010).

Finally, in a forecasting exercise we find that the QBLL estimators deliver excellent forecast performance. Their ability to accommodate drifting parameters nonparametrically delivers forecast improvements over fixed parameter models, such as virtually unbiased forecasts and uniform PITs. In addition, their capacity to considerably increase the number of variables can improve both point and density forecasts of small-dimensional models. We find that both drifts in the volatility and in the parameters can have important impact on forecasting: drifts in the volatility improve density forecasts while time variation in the autoregressive component significantly reduces forecast bias.

3 Time Varying Parameter DSGE Model

3.1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models are popular tools extensively used in both academic work and macroeconomic policy making. Their success is a result of their capacity to combine economic microfoundations derived from optimisation decisions of agents with rational expectations and business cycle fluctuations. Traditionally, the consensus in the macroeconomic literature has been that there exists an apparent trade-off between theoretical coherence, whereby a model's outcomes can be explained by well-established theory, and empirical coherence, whereby a model can fit and explain macroeconomic data well, but its outcomes are often difficult to interpret or justify from a theoretical standpoint. Models that exhibit theoretical and empirical coherence simultaneously were deemed infeasible. DSGE models were alleged to be at the theoretical end of this trade-off curve. On the other hand, reduced-form models, such as VAR models, exploiting correlations in time series with little reliance on macroeconomic theory, were put at the empirical end. It was the work of Smets and Wouters (2003, 2005, 2007), based on earlier work of Rotemberg and Woodford (1997) and Christiano, Eichenbaum and Evans (2005), that changed this perception and demonstrated that medium-sized DSGE models can be successfully taken to the data and produce superior forecasts to standard BVAR models. Following Smets and Wouters, the literature on DSGE model estimation and forecasting has become a vibrant area of research with considerable progress in the development of the underlying economic theory and the design of numerical solution and estimation algorithms.

At the heart of DSGE models are so called deep parameters that define the preferences and technological environment of the economy. These are kept constant and are structural in the sense that they are not subject to the Lucas critique - they are invariant to both policy and structural shocks. There are two issues related to these parameters that this Chapter will address. First, it is important to recognise the possibility of parameter drift in order to re-evaluate the usefulness and relevance of DSGE models. If substantial evidence is found that some of these structural parameters are in fact not constant, this could be interpreted as a need to revise existing models in order to account for such variation. It is possible that slow time variation is the outcome of long term cultural or technological shifts in the economy that DSGE models are ill-equipped to model, since they focus primarily on business cycle fluctuations. Nevertheless, taking into account such slow variation is paramount for the effective use of DSGE models. Furthermore, time variation in these parameters can be a signal for misspecification in existing models and hence a guide to amend and improve them. Second, these models are widely used in forecasting both by academics and official institutions. Hence, allowing the structural parameters to change and using only their most recent values for generating predictions seems like a useful modification that would be expected to improve forecasting performance, possibly at the cost of making the separation between structural and reduced-form models less clear.

To accommodate such time variation in DSGE model parameters, this Chapter applies a quasi-Bayesian Local Likelihood (QBLL) method, developed in a general reduced-form setting in Chapter 2. QBLL estimates parameters at each point in time, appropriately weighting the sum of log likelihoods of the sample, with weights generated by a kernel function. The method is general and can be applied to any DSGE model. Furthermore, for generating forecasts, it is no more computationally intensive than estimating a DSGE model with fixed parameters.

This and next two Chapters contribute to a small but expanding literature on estimating DSGE models with time variation in the parameters which has two strands. Fernandez-Villaverde and Rubio-Ramirez (2008) and Justiniano and Primiceri (2008) model time variation by assuming stochastic processes for a subset of the parameters and include these to the set of state equations. For instance, Fernandez-Villaverde and Rubio-Ramirez (2008) assume that the agents, in the model, take into account current and future parameter variation, utilising the parameters' representation as stochastic processes when computing their expectations. A similar assumption is made by Schorfheide (2005), Bianchi (2013), Foerster, Rubio-Ramirez, Waggoner and Zha (2014), but the parameters are modelled as Markov-switching processes. In contrast, Canova (2006), Canova and Sala (2009) and Giacomini and Rossi (2009) assess parameter time variation by estimating DSGE models over rolling samples. A similar strategy was followed by Castelnuovo (2012), Cantore, Levine and Melina (2012) and Canova and Ferroni (2012). It is useful to contrast our work with both these strands. This first strand makes parametric assumptions about the variation in the parameters. These assumptions are not microfounded but have a reduced form flavour. Instead, the method presented in the current Chapter is agnostic about the source of the variation apart from assuming that it is slow, although given some time it can track more abrupt forms of change. Given the considerable likelihood that any changes are the result of long term cultural and technological shifts that no mainstream business cycle model is well equipped to explain, this agnostic approach has merit. A further issue is that computational complexity restricts the ability of allowing for time variation to only a small subset of the model parameters whereas our approach is scaleable to the full set of parameters. Concerning the second strand, our proposed method employs a nonparametric kernel-based procedure that encompasses rolling window estimation as a special case. The approach presented in the current Chapter is an extension and formalisation of rolling window estimation, generalised by combining kernel-generated local likelihoods with appropriately chosen priors to generate a sequence of quasi-posterior distributions for the objects of interest over time, following the methodology developed in Chapter 2. Evidence provided in Giraitis et al. (2014) suggests that other kernel functions may have more desirable propeties than the flat kernel underlying the rolling window. Our approach is related to the one in Giraitis, Kapetanios, Theodoridis and Yates (2013), but they apply the local kernel estimator developed by Giraitis et al. (2014) to the minimum distance estimator that matches DSGE and VAR impulse responses, to provide a frequentist estimation approach. Both the kernel and the rolling window approaches, when applied to structural models, assume that, instead of being endowed with perfect knowledge about the economy's data generating process, agents take parameter variation as exogenous when forming their expectations about the future. This assumption facilitates estimation and can be rationalised from the perspective of models featuring learning problems, where agents form beliefs about the parameters based on observing past data. For example, Cogley and Sargent (2009) utilise Kreps (1998)'s anticipated utility approach, where in each period agents employ their current beliefs as the true (time invariant) parameters. They show that in the presence of parameter uncertainty, the anticipated utility approach outperforms the rational expectation approximation. A recent application of the anticipated utility approach is Johannes, Lochstoer and Mou (2015), where assets are priced at each point in time, using current posterior means for the parameters and assuming that current values will last indefinitely in the future. At each period, agents learn the new parameter values and adjust their expectations[20].

One aim of the following Chapters is to improve the accuracy of DSGE models in forecasting. Smets and Wouters (2007) show that their medium-sized DSGE model can generate forecasts for seven US macro variables that are superior to those obtained from a BVAR model. The gains of the structural model over the reduced-form model are substantial especially at longer horizons. Additional evidence that DSGE models may deliver competitive forecasts in comparison with statistical models and survey of professional forecasters is provided by Rubaszek and Skrzypczynski (2008), Steinbach, Mathuloe and Smith (2009), Edge, Kiley and Laforte (2009), Edge and Guerkaynak

[20] A similar outcome can be achieved in a linearised DSGE model with random walk processes for the drifting parameters (a frequently made assumption in the time varying parameter VAR literature), where rational expectations on the side of agents would imply that the future values of the parameters are equal to the current posterior means.

(2010), Wieland and Wolters (2011) and Del Negro and Schorfheide (2013b). To the best of our knowledge, there is no documented evidence of the forecasting performance of a DSGE model with time variation in the parameters. The closest to ours is the working paper by Edge, Guerkaynak and Kisacikoglu (2013) who use rolling window scheme to assess the forecasting record of a DSGE model.

The Chapter is organised as follows. Section 3.3 introduces the Quasi-Bayesian Local Likelihood approach in the context of DSGE models including a Metropolis within Gibbs algorithm for heteroscedastic time invariant DSGE model, Section 3.4, 3.5 and 3.6 present an empirical application based on the Smets and Wouters (2007) model, Section 3.7 provides a forecasting comparison and Section 3.8 concludes.

3.2 Time Variation in DSGE Models

In the context of DSGE model estimation, there are advantages of adopting a Bayesian approach. Bayesian methods provide a natural way of combining econometric estimation with information provided by calibration methods widely used in the previous generation of models (see Kydland and Prescott (1996)). For example, by construction, we know that the discount factor, β, that consumers use to discount expected future utility cannot take negative values, is bounded between zero and one and a typical value based on the assumption of a 4% annual discount rate is 0.99. Adding a probability mass in the form of a prior is a natural way to incorporate such additional information which is not contained in the data and serves as augmenting the likelihood with artificial observations. In addition, the likelihood of DSGE models may often be ill-identified or not globally concave. Adding a prior can resolve such issues and make the problem well-defined (see, Lindley (1971)). Finally, a Bayesian approach can deal in a natural way with model misspecification. Instead of assuming that there is a unique parameter vector that contains the "true" values of all parameters, the Bayesian approach considers the parameters as random variables and the estimation procedure as a learning process with respect to the characteristics of these random variables, after incorporating information on the available data. An and Schorfheide (2007) and Del Negro and Schorfheide (2013a) offer a detailed review of Bayesian inference in the context of DSGE models.

Here, we follow the methodology from Chapter 2. Let $p_t(\theta_t)$ denote a prior distribution for θ_t at time t. Then, the quasi-posterior $p_t(\theta_t|Y)$ is given by:

$$p_t(\theta_t|Y) = \frac{p_t(\theta_t)L_t(Y|\theta_t)}{\int_\Theta p_t(\theta_t)L_t(Y|\theta_t)d\theta} \propto p_t(\theta_t)L_t(Y|\theta_t)$$

where $L_t(Y|\theta_t)$ is the local likelihood function, $L_t(Y|\theta_t) = \prod_{j=1}^{T} L(y_j|y^{j-1}, \theta_t)^{w_{tj}}$ for $t = 1, .., T$, $Y = (y_1, .., y_T)'$, and $L(y_j|y^{j-1}, \theta_t)$ denotes the likelihood for observation j, conditional on the history y^{j-1}. As demonstrated in Chapter 2, this provides a generic quasi-Bayesian principle for estimating general time varying coefficient models that require little more than standard Bayesian numerical techniques applicable to fixed coefficient models. The discussion and results in Chapter 2 related chiefly to time variation in reduced form models. One advantage of our method is that it is applicable without conjugacy: if the posterior did not belong to a known distributional family, other MCMC methods can be used to generate draws from that posterior. Here, we extend the method and outline how it can be applied to the estimation of a DSGE model.

The literature on time varying DSGE models is less developed than that on time varying reduced form models and more contraversial. The alternative approach of specifying processes for the drifting parameters in a DSGE context has been applied for instance by Fernandez-Villaverde and Rubio-Ramirez (2008). The main advantage of their approach is that agents populating the model take into account the parameters' stochastic processes when forming their expectations about the future. Our econometric approach, on the other hand, does not incorporate a law of motion for the parameters when solving the agents' rational expectation problem. However, if the parameters of the model are driven by either a time varying deterministic or slowly moving stochastic process, as the popular random walk assumption in the literature, then our econometric approach does not violate the rational expectation assumption in a linearised model because future changes in the parameters are unpredictable by both agents and the econometrician. This implies that current parameter values are the best prediction for future values anyway. Another disadvantage of modelling parameters' variation by explicitly specifying a stochastic process is that it is subject to the curse of dimensionality. The state vector needs to be augmented for each parameter allowed to vary and an additional shock is introduced. Because of this dimensionality problem, all parameters cannot be modelled simultaneously in this way. For instance, Fernandez-Villaverde and Rubio-Ramirez (2008) do not allow both Taylor rule and price rigidity parameters to vary simultaneously when estimating their DSGE model. Our alternative econometric approach does not suffer from such dimensionality issues.

In addition, the modelling approach of Fernandez-Villaverde and Rubio-Ramirez (2008) imposes an additional structure by relying on the assumption that the law of motion for the parameters' time variation is correctly specified. Our nonparametric approach performs well for many different parameters' laws of motion. Chapter 2 demonstrated in a Monte Carlo exercise that if the law of motion is misspecified, inconsistent estimates of the parameters' time variation are obtained if they are treated as unobserved state variables as in Fernandez-Villaverde and Rubio-Ramirez (2008). In contrast, the results in Chapter 2 suggest that our nonparametric alternative is consistent for a wide class of parameter processes. Schorfheide (2007) argues that by treating time varying parameters as unobserved state variables as in Fernandez-Villaverde and Rubio-Ramirez (2008), identification issues which are attenuated by the use of priors, as argued earlier in this Section, may arise[21]. Our approach has the advantage of being able to incorporate prior information about the time varying parameters at each point in time to solve possible identification issues.

3.3 The Quasi-Bayesian Local Likelihood Method for DSGE Models

In this Section, we show how to apply the QBLL approach described in Chapter 2 to a DSGE model with linear state-space representation. Note, however, that the QBLL method could also be applied to models that have a non-linear state space representation such as in Fernandez-Villaverde and Rubio-Ramirez (2007).

The linearized rational expectation model with time varying parameters can be written in the form:

$$A(\theta_t)\mathbb{E}_t x_{t+1} = B(\theta_t)x_t + C(\theta_t)v_t, \qquad v_t \sim N(0, Q(\theta_t))$$

where x_t is a $n \times 1$ the vector of model's variables, v_t is a $k \times 1$ vector of structural shocks, θ_t is a vector of parameters, including parameters governing preferences and the shocks' stochastic processes, A, B and C are matrices, which are functions of θ_t, $Q(\theta_t)$ is a diagonal covariance matrix, and E_t is the expectation operator conditional on information available at time t. Observe that we have one such equation at each point in time $t = 1, ..., T$.

A numerical solution of the rational expectation model can be obtained by one of the available methods (for instance, Blanchard and Kahn (1980) or Sims (2002)). The resulting state equation is given by:

$$x_t = F(\theta_t)x_{t-1} + G(\theta_t)v_t \tag{34}$$

where the $n \times n$ matrix F and $n \times k$ matrix G can be computed numerically for a given parameter

[21] For instance, Fernandez-Villaverde and Rubio-Ramirez (2008) obtain values of their Taylor rule parameters for which the Taylor principle is not satisfied.

vector θ_t. The system is augmented with a measurement equation:

$$y_t = D(\theta_t) + Z(\theta_t)x_t \tag{35}$$

where y_t is an $m \times 1$ vector of observables, typically of a smaller dimension than x_t (i.e. $m < n$) and Z is a $m \times n$ matrix that links those observables to the latent variables in the model x_t.

Equations (34) and (35) provide the state-space representation of the model, which is linear and Gaussian at each successive set of parameters θ_t for $t = 1, ..., T$. Therefore, the Kalman filter can be employed to recursively build the likelihood of the sample of observables $\{y_t\}_{t=1}^T$. The appropriately weighted likelihood of the sample is given by:

$$L_t(Y|\theta_t) = \prod_{j=1}^{T} L(y_j|y^{j-1}, \theta_t)^{w_{tj}} \text{ for } t = 1, .., T$$

where w_{tj} is an element of the $T \times T$ weighting matrix $W = [w_{tj}]_{t,j=1}^T$, computed using a kernel function given in equation (4):

$$w_{jt} = \tilde{w}_{jt} / \sum\nolimits_{t=1}^{T} \tilde{w}_{jt}, \quad \tilde{w}_{jt} = K\left(\frac{j-t}{H}\right) \quad \text{for} \quad j, t \in \{1, ..., T\}.$$

with a bandwidth H, satisfying the conditions in Chapter 2.

In the fixed parameter case, the weights on each likelihood sum up to T. In our case, each row of W is normalised to sum up to $2H + 1$, such that:

$$\sum_{j=1}^{T} w_{tj} = 2H + 1 \qquad t = 1, ..., T.$$

This normalisation is employed in order to maintain the relative weights between the likelihood and the prior[22].

For the application presented in this Chapter, the Normal kernel function in (5) is used to generate the weights. If the bandwidth H goes to infinity, the likelihood would collapse to the fixed parameter case, where each likelihood is weighted equally. If H is small, the weights are concentrated around a single observation. Our choice of bandwidth is $H = T^{0.5}$, motivated by the

[22] This chapter contains earlier work in which we had an intuition that the sum of the weights needs to diverge with the sample size to obtain consistency; we had not, however, obtained the result in Proposition 1 yet, which gives us the normalising rate. Hence, here and in the following two chapters, we normalise the weights to sum to $2H + 1$, which results in a consistent estimator with variance larger than the asymptotic variance of the estimator in chapter 2, where the weights sum up to $\varkappa_{jT}$.

optimal bandwidth choice used for inference in time varying random coefficient models (see Giraitis et al. (2014)).

The local likelihood of the DSGE model at point t, denoted $L_t(Y|\theta_t)$, is augmented with the prior distributions for the parameters, $p_t(\theta_t)$, to get the quasi-posterior at time t, $p_t(\theta_t|Y)$:

$$p_t(\theta_t|Y) = \frac{L_t(Y|\theta_t)p_t(\theta_t)}{p(Y)} \propto \prod_{j=1}^{T} L(y_j|y_{j-1},\theta_t)^{w_{tj}} p_t(\theta_t).$$

It should be noted that, for our DSGE applications, we assume the prior $p_t(\theta_t)$ to be fixed over time, i.e., $p_t(\theta_t) = p(\theta_t)$ for all t. One could potentially allow the prior to be time varying, exploring the idea that the posterior yesterday can be used for a prior today. However, since we would like to explore only the possibility of parameter drift, we choose to be agnostic about time variation in the parameters before the estimation and keep the prior values fixed over time.

3.3.1 Characterising the Posterior Distributions

To obtain the joint posterior distribution of the parameters, we need numerical methods because the matrices F and G are non-linear functions of θ, and hence the posterior does not fall in known families of distributions with moments that could be derived analytically. The most commonly used procedure to generate draws from the posterior distribution of θ is the Metropolis-Hastings (MH) algorithm, proposed by Metropolis et al. (1953) and generalised by Hastings (1970). Although the quasi-posterior distribution could be obtained by other methods, such as the Importance Sampling (IS) algorithm, the MH algorithm delivers good convergence under fairly general regularity condition (see Geweke (1999, 2005)) and asymptotically normal posterior distribution (see Walker (1969), Crowder (1988) and Kim (1998)).

The algorithm described here is version of Schorfheide (2000)'s Random Walk Metropolis (RWM) algorithm, modified to include the kernel weighting scheme. Our aim is to obtain a sequence of posterior distributions $p_t(\theta_t|Y)$ for each point in time $t = 1, .., T$. At each t the algorithm implements the following steps.

Step 1. The posterior is log-linearised and passed to a numerical optimisation routine. Optimisation with respect to θ is performed to obtain the posterior mode:

$$\widehat{\theta}_t = \arg\min_{\theta} \left(-\sum_{j=1}^{T} w_{tj} \log L(y_j|y^{j-1},\theta_t) - \log p(\theta_t) \right).$$

Step 2. Numerically compute $\widehat{\Sigma}_t$, the inverse of the (negative) Hessian, evaluated at the posterior mode, $\widehat{\theta}_t$.

Step 3. Draw an initial value θ_t^0 from $N(\widehat{\theta}_t, c_0^2\widehat{\Sigma}_t)$.

Step 4. For $i = 1, ..., n_{sim}$, draw ζ_t from proposal distribution $N(\theta_t^{(i-1)}, c^2\widehat{\Sigma}_t)$. Compute

$$r(\theta_t^{i-1}, \zeta_t|Y_{1:T}) = \prod_{j=1}^{T} L(y_j|y^{j-1}, \zeta_t)^{w_{tj}} p(\zeta_t) / \prod_{j=1}^{T} L(y_j|y^{j-1}, \theta_t^{i-1})^{w_{tj}} p(\theta_t^{i-1}),$$

which is the ratio between the weighted posterior at the proposal ζ_j and θ_t^{i-1}.

The draw ζ_t is accepted (setting $\theta_t^i = \zeta_t$) with probability $\tau_t^i = min\{1, r(\theta_t^{(i-1)}, \zeta_t|y_{1:T})\}$ and rejected ($\theta_t^i = \theta_t^{i-1}$) with probability $1 - \tau_t^i$. c_0^2 and c^2 are scaling parameters adjusting the step size of the MH algorithm in order to get desirable rejection rates such that we achieve convergence. The literature supports setting the scaling parameters such that acceptance rates of between 20% and 40% are achieved[23].

3.3.2 Computing Forecasts

Once the time varying quasi-posterior distribution of the parameters is obtained using the algorithm in Section 3.3.1, we can compute out-of-sample forecasts for the observables y. For this task, we only need the corresponding quasi-posterior distribution at the end of each in-sample, $p(\theta_{t=T}|Y)$, which contains the most recent values of the model's parameters and hence the most relevant information for predicting the future. Therefore, for generating DSGE-based predictions, our method is as computationally intensive as forecasting with standard fixed parameter DSGE models: it requires the computation of the posterior only once.

The predictive distribution of the sample $p(y_{T+h}|y_{1:T})$, h horizons ahead, is given by the conditional probability of the forecasts, averaged over all possible values of the parameters, the unobservables at the end of the sample x_T, and all possible future paths of the unobservables $x_{T+1:T+h}$:
$p(y_{T+h}|y_{1:T}) =$

$$\int_{(x_T,\theta_T)} \left(\int_{x_{T+h}} p(y_{T+h}|x_{T+h}) p(x_{T+h}|x_T, \theta_T, y_{1:T}) dx_{T+h} \right) p(x_T|\theta_T, y_{1:T}) p(\theta_T|y_{1:T}) d(x_T, \theta_T)$$

[23] In particular, Roberts, Gelman and Gilks (1997) show that, under some conditions, the optimal asymptotic acceptance rate is 23.4%.

where $p(\theta_T|y_{1:T})$ is the posterior of the parameters at the end point of the in-sample period, T. We employ a slightly modified version of the algorithm for generating draws from the predictive distribution outlined in Del Negro and Schorfheide (2013b). The algorithm is as follows.

Step 1. Using the saved draws from the posterior at the end of the sample $p(\theta_T|y_{1:T})$, for every draw $k = 1, .., n_{sim}$ (or for every n-th draw if thinning is required), apply the Kalman filter to compute the moments of the unobserved variables at T using the density $p(x_T|\theta_T^k, y_{1:T})$.

Step 2. Draw a sequence of shocks $v_{T+1:T+h}^k$ from a $N(0, Q(\theta_T^k))$, where $Q(\theta_T^k)$ is a draw from the quasi-posterior distribution of the diagonal variance-covariance matrix of the shocks at T. For each draw k from $p(\theta_T|y_{1:T})$ and from $p(x_T|\theta_T^k, y_{1:T})$, use the state equation to obtain forecasts for the unobserved variables:

$$\widehat{x}_{T+1:T+h}^k = F(\theta_T^k)x_{T:T+h-1}^k + G(\theta_T^k)v_{T+1:T+h}^k.$$

Step 3. Use the forecast simulations for the latent variables in the measurement equation:

$$\widehat{y}_{T+1:T+h}^k = D(\theta_T^k) + Z(\theta_T^k)\widehat{x}_{T+1:T+h}^k.$$

Using the above algorithm, we obtain a predictive density of n_{sim} draws of $\widehat{y}_{T+1:T+h}^k$ which can be used to derive numerical approximations of moments, quantiles and densities of the out-of-sample forecasts.Finally, point forecasts are obtained by computing the mean of the distribution of $\widehat{y}_{T+1:T+h}^k$ for each forecasting horizon.

3.3.3 Nonparametric Heteroscedasticity in a DSGE Model

In the previous subsection, we show how to apply the QBLL estimation procedure to a linearised DSGE model in order to estimate all the model's parameters varying over time. The idea of the algorithms outlined in Chapter 2 is general and can be applied in non-linear and non-conjugate setups such as DSGE models. Here we outline a Metropolis within Gibbs algorithm that can be used to consider time variation in the variances of the DSGE exogenous shocks while keeping the structural parameters that characterise the preferences of the agents in the model fixed. We use the structure of the algorithm presented in Justiniano and Primiceri (2008), modified to include the quasi-Bayesian local likelihood estimator of the estimation of the time varying covariance matrix.

The linearized rational expectation model with heteroscedastic disturbances can be written in the form:

$$A(\theta)\mathbb{E}_t x_{t+1} = B(\theta)x_t + C(\theta)\eta_t, \qquad \eta_t \sim N(0, \Sigma_t)$$

where x_t is a $n \times 1$ the vector of model's variables, v_t is a $k \times 1$ vector of structural shocks, θ is a vector of parameters, A, B and C are matrices, functions of θ, Σ_t is a diagonal time-varying covariance matrix, and $\mathbb{E}_t$ is the expectation operator conditional on information available at time t.

A numerical solution of the rational expectation model can be obtained with a resulting state equation is given by:

$$x_t = F(\theta)x_{t-1} + G(\theta)\Sigma_t^{1/2}\vartheta_t, \qquad \vartheta_t \sim N(0, I_n) \tag{36}$$

where the $n \times n$ matrix F and $n \times k$ matrix G can be computed numerically for a given θ. Note that because the DSGE model is linearised, the solution in (36) does not depend on the history of $\Sigma_{1:T}$.The system is augmented with a measurement equation:

$$y_t = D(\theta) + Z(\theta)x_t \tag{37}$$

where y_t is an $m \times 1$ vector of observables, typically of a smaller dimension than x_t (i.e. $m < n$) and Z is a $m \times n$ matrix that links those observables to the latent variables in the model x_t.

Conditional on the history of $\Sigma_{1:T}$, equations 36 and 37 are a linear Gaussian state-space system; so standard Kalman filter techniques can be employed to recursively build the likelihood of the sample of observables $\{y_t\}_{t=1}^T$. The likelihood of the sample, combined with a prior distribution for the parameter vector θ, is given by:

$$p(\theta|Y) \propto \prod_{t=1}^{T} L(y_t|y^{t-1}, \theta)p(\theta).$$

To draw from the posterior, a Metropolis step will be employed. Conditional on the hiatory $\Sigma_{1:T}$ and on the Metropolis draw from the posterior of θ, a draw from the structural shocks v_t can be obtained using Carter and Kohn (1994) or Durbin and Koopman (2002)'s disturbance smoothers.

Conditonal on a draw of the history of the shocks $v_{1:T}$ and on the Metropolis draw of θ, we have that: $v_t = \Sigma_t\eta_t$. This means we are in the setting of Proposition 7 so by specifying a Gamma[24]

[24] Note that since Σ_t is a diagonal matrix, the Wishart prior in Proposition 7 $W(\alpha_{0t}, \gamma_{0t})$ can be written as Gamma prior for the i^{th} diagonal element $Ga(\alpha_{0t}^i/2, \gamma_{0t}^i/2)$. This makes a difference computationally, as it is faster to draw

prior for each diagonal element of Σ_t^{-1} with parameters $\alpha_{0t}^i/2, \gamma_{0t}^i/2$, the posterior at each point in time t is also Gamma with parameters:

$$\tilde{\alpha}_t^i/2 = \alpha_{0t}^i/2 + \sum_{j=1}^{T} w_{tj}/2 \tag{38}$$
$$\tilde{\gamma}_t^i/2 = \gamma_{0t}^i/2 + \frac{1}{2}V' D_t V$$

where $V = [v_{1,...,} v_T]'$ and $D_t = diag(w_{tj})$ and w_{tj} is the t, j^{th} element of the kernel weighting matrix defined in (4).

Algorithm Outline: Metropolis within Gibbs Suppose we are in iteration g, so we have $\theta^g, H^{T,g}, \Phi^g, \widehat{\eta}_t^g, m^{Tg}$:

Then iteration $g+1$ looks like this:

Step 1. Draw the history of structural shocks $\widehat{\eta}_{1:T}^{g+1}$ using Carter and Kohn (1994) or Durbin and Koopman (2002) algorithms the state space below.

$$x_t^g = F(\theta^g)x_{t-1}^g + G(\theta^g)\widehat{\eta}_t$$

$$y_t = D(\theta^g) + Z(\theta^g)x_t^g$$

Step 2. Draw the volatilities $\Sigma_{1:T}^{g+1}$ using the novel QBLL approach from an inverse - Gamma[25] at each point in time t with parameters given in (38).

Step 3 (Metropolis Step). Draw θ^{g+1}, conditional on the draw from the history of volatilities $\Sigma_{1:T}^{g+1}$. In particular, draw ϑ from the proposal distribution $N(\theta^g, c^2\widehat{\Lambda})$, where $\widehat{\Lambda}$ is the Hessian evaluated at the posterior mode and c^2 is a scaling parameter that controls the step size through the parameter space (and hence the rejection rate of the Metropolis).

Compute

$$r = \prod_{t=1}^{T} L(x_t|\vartheta)p(\vartheta) / \prod_{t=1}^{T} L(x_t|\theta^g)p(\theta^g)$$

Accept $(\theta^{g+1} = \vartheta)$ with probability $\tau = min\{1, r\}$ and reject $(\theta^{g+1} = \theta^g)$ with probability $1-\tau$.

from univariate distrbutions.

[25] This is the case as if $\sigma^{-2} \sim Ga(\alpha, \gamma)$, then $\sigma^2 \sim Inv - Ga(\alpha, \gamma)$.

3.4 Model and Data

The DSGE model to which we apply our quasi-Bayesian Local Likelihood approach is the model from Smets and Wouters (2007), which is an extension of a small-scale monetary RBC model with sticky prices (such as Goodfriend and King (1997), Rotemberg and Woodford (1997), Woodford (2003), Ireland (2004) and Christiano et al. (2005)). In addition to the sticky prices, the model also contains some additional shocks and frictions, including sticky nominal price and wage settings with backward inflation indexation, investment adjustment costs, fixed costs in production, habit formation in consumption and capital utilization. It also features seven exogenous shocks that drive the stochastic dynamics of the model. The foundations of the model are derived from the intertemporal optimisation problems of different agents. In particular, there are seven types of agents in the model: consumers that supply labour, choose consumption level, hold bonds and make investment decisions; intermediate goods producers which are in a monopolistically competitive market and cannot adjust prices at each period and final goods producers, who buy intermediate goods, package them and resell them to consumers in a perfectly competitive market. In addition, there is a labour market with a similar structure: there are labour unions with market power that buy the homogenous labour from households, differentiate it, set wages and sell it to the labour packers, who package it and resell it to intermediate goods producers in a perfectly competitive environment. Finally, there is a central bank that follows a nominal interest rate rule, adjusting the policy instrument in response to deviations of inflation or output from their target levels and a government that collects lump-sum taxes which appear in the consumer's budget constraint and whose spending is exogenously driven.

The model is log-linearised around its steady state and trended variables are detrended with a deterministic trend[26]. The model is estimated using seven macroeconomic quarterly time series for the United States for the period of 1964Q3 to 2012Q4 as observables. The variables are the ones used in Smets and Wouters (2007), namely, output, consumption, investment and wages per capita growth; inflation, hours and the interest rate (see Appendix 7.2 for more details).

In this Section we present results for the fixed parameter model and also for the version with time varying parameters estimated with QBLL method described in the previous Section. In both cases, we employ the priors from Smets and Wouters (2007), with a number of MH draws of $220,000$,

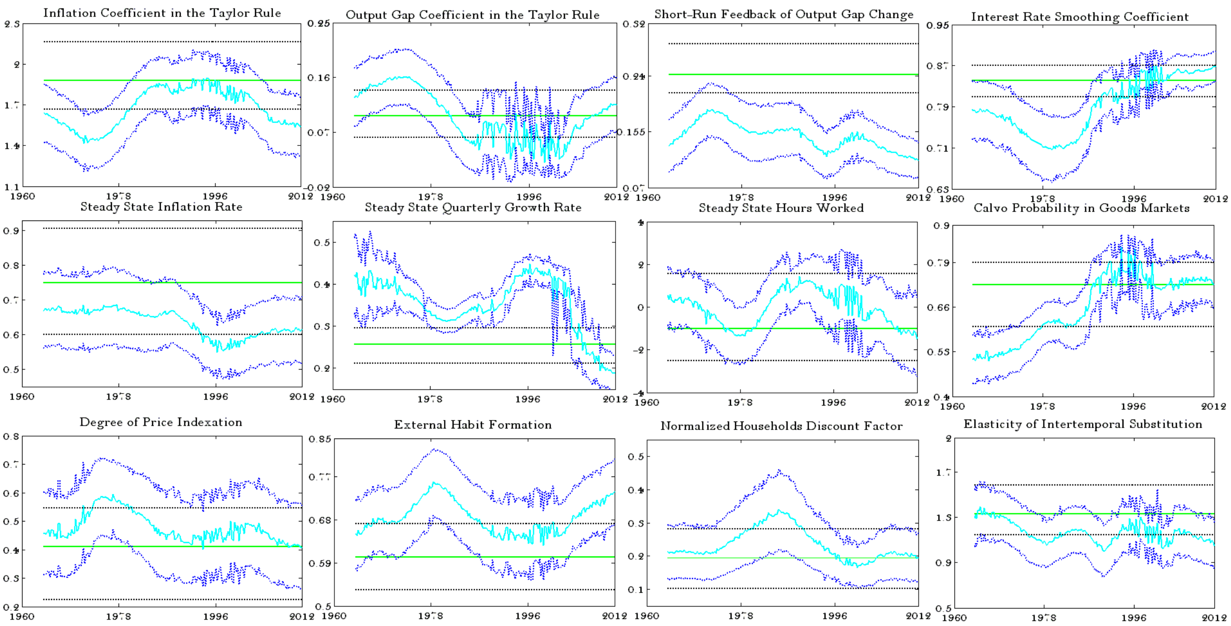
[26] The linearised model is presented in the Appendix. For full derivations from the non-linear first order conditons, please refer to the Technical Appendix in Smets and Wouters (2007) available at: http://www.aeaweb.org/aer/data/june07/20041254_app.pdf

from which we drop the first 20, 000. We set the scaling parameters such that acceptance rates are around 25%. We apply the QBLL method using the Normal kernel function in equation (5) with a bandwidth size of $T^{0.5}$.

3.5 Results

In this Section, we discuss the parameter estimates (Figures 11-13). We employ Figures 11-13 to judge informally whether a parameter's variation is substantial by checking whether the QBLL estimates are outside the confidence bands of the fixed-parameter estimates. We adopt Fernandez-Villaverde and Rubio-Ramirez (2008)'s definition of 'structural' parameters: these are preference and technology parameters which are invariant to both policy and shocks. If a parameter is found to be within the fixed-parameter 68% bands, we conclude that it is in fact 'structural'. If a parameter varies smoothly over time, following a clear pattern, we infer that it has been a subject to structural change. On the other hand, if a parameter exhibits an erratic time variation, we would point to a possible misspecification of that parameter. The solid blue line are the time varying estimates obtained with the QBLL method and 68% confidence bands are represented by the dotted blue lines. The green line represents the full sample fixed-parameter estimates, with the dotted lines around it - 68% confidence bands.

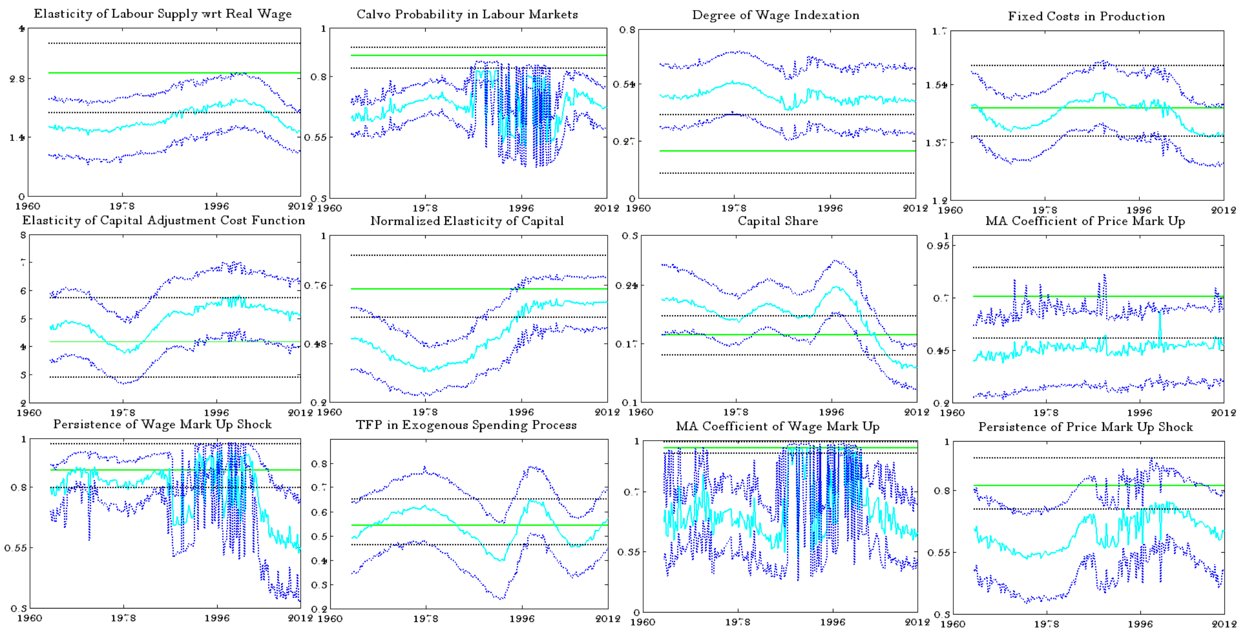
Figure 11: The DSGE parameters over time



The top panel of Figure 11 assesses results for the policy preference parameters. Our results are broadly consistent with previous studies (Clarida et al. (2000), Cogley and Sargent (2002), Fernandez-Villaverde and Rubio-Ramirez (2008)) that found evidence of structural changes in Taylor rule parameters. The Federal Reserve has shifted its policy priority from output towards inflation since the parameter that measures the reaction to inflation increases between 1979 and 1996, while the reaction to the output gap decreases over this period. The interest rate smoothing parameter is lower in the 1980s than in later periods, while the steady state inflation rate decreases between 1985 and 1996.

The second panel of Figure 11 provides evidence of changes in the steady-state growth rate of per capita output (as well as consumption, investment and the real wage, which share the same trend). During most of the period and up to 2005, the QBLL posterior mean for this parameter is around 0.4, that is, an annual growth rate of 1.6%, however, this decreases to 1% annually in period of the 2007-8 financial crisis. In contrast, fixed-parameter estimates under-estimate these values over most of the sample, but over-estimate it during the recent period. This parameter is important for generating forecasts as it appears in several of the measurement equations and Kolasa and Rubaszek (2015) bring attention to the importance of this parameter in reducing forecasting bias.

Figure 12: The DSGE parameters over time



The findings on the price rigidity parameters are consistent with the evidence presented in Fernandez-Villaverde and Rubio-Ramirez (2008). In particular, we document a negative relation between the price indexation (last panel in Figure 11) and price stickiness (second panel in Figure 11) parameters after the mid-1970s; hence, periods characterised by high Calvo probability parameter are also of low indexation and vice versa. The fall in inflation indexation during the Great Moderation is consistent with findings of Gali and Gertler (1999) and could be explained by the decreased need to adjust prices frequently due to low and stable inflation leading to longer length of contracts[27] and hence higher Calvo probability parameter and lower indexation. The variation uncovered in the price rigidity parameters suggests that there is no stable predictive relation between inflation and output gap over time (i.e. the Phillips curve has become flatter in the low volatility period of the Great Moderation) which cast doubt on the ability of Calvo pricing models to adequately capture pricing behavior of firms and unions in the economy.

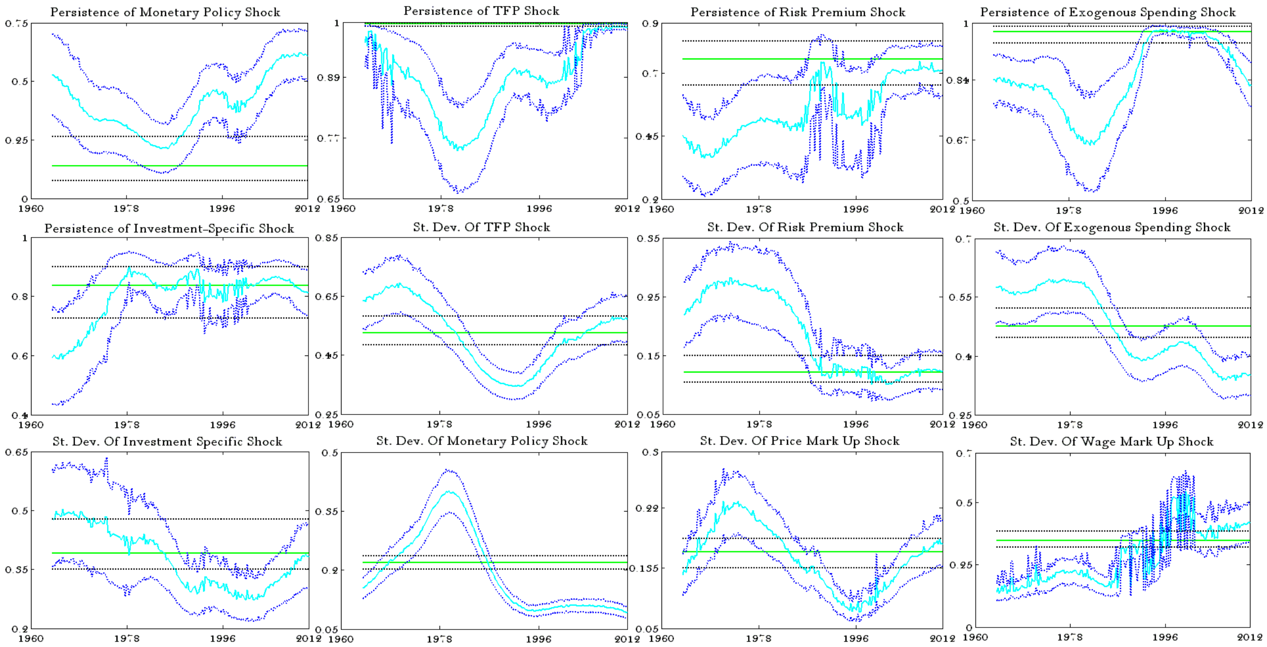
There are parameters that appear to move very little over the entire sample period or seem to remain within the confidence bands of the fixed parameter estimates throughout most of the sample, such as the elasticity of intertemporal substitution and the household discount factor (last panel of Figure 11), and the elasticity of labour supply or the fixed production costs (first panel of Figure 2). We draw comfort on those results, as they could be interpreted as evidence of the structural nature of these parameters.

On the other hand, the moving average (MA) coefficient, the persistence coefficient (last panel of Figure 12) and the standard deviation (last panel in Figure 13) of the wage mark-up shock process, as well as the Calvo parameter in labour markets (top panel in Figure 12) all appear very volatile and this could be evidence that they are seriously misspecified and should not be kept fixed. Interestingly enough, all four are parameters that govern labour market dynamics through the wage equation and all become very unstable during the Great Moderation period. This could be interpreted as evidence that during the Great Moderation, a Calvo model with an ARMA wage shock may not have been an adequate model to characterise the dynamics of the labour market in the US. An alternative interpretation is that there might be insufficient information in the data in order to jointly identify all four parameters during the Great Moderation.

The standard deviations of the structural shocks (panel 2 and 3 of Figure 13) also move in the expected direction, consistent with findings of low stochastic volatility during the Great Moderation (e.g. Primiceri (2005) and Sims and Zha (2006)). In particular, all shocks' volatilities fall in the late 1980s and remain low throughout the 1990s. Moreover, their posterior distributions are narrower during that period implying that there is less uncertainty about the possible values they

[27] The average price duration is given by $\frac{1}{1-\xi_p}$, where ξ_p is the Calvo probability in the goods market.

Figure 13: The DSGE parameters over time



can take. The standard deviation of the monetary policy shock, for instance, peaks in the 1980s, implying a larger role of the shock throughout that period and falls considerably after the 1990s, having lesser impact on the business cycle as a consequence of the more adequate policy. For some shocks' standard deviations (e.g. TFP, investment-specific technology and price-mark up shocks) we observe an increase in the end of the sample leading to the recent financial crisis. Due to the considerable time variation we uncover in the volatilities of the shocks, using the most recent values of the estimated volatility parameters when generating forecasts is expected to improve the density of the forecasts compared to simply using the fixed parameter estimates that average these over the entire in-sample period. Finally, the autoregressive coefficients for the stochastic processes (panels 1 and 2 of Figure 13) seem to move considerably, which is unsurprising as these are designed to capture dynamics in the data. They are not truly structural, in the sense that there is no underlying macroeconomic theory that implies that they are not subjected to shocks or policy. Most of the shocks' persistence coefficients display a U-shape with low persistence towards the end of the Oil Crises and higher persistence during the Great Moderation. For instance, the TFP shock becomes very persistent during the recent crisis with AR coefficient very close to one, implying almost permanent shock to productivity.

3.6 Time Varying Impulse Response Functions

In this Section, we turn to the estimated impulse response functions over time. We investigate whether there is evidence for structural change in the transmission mechanism of important variables to macroeconomic shocks, resulting from the documented time variation in the paramerers.

Figure 14 displays the impulse response functions of output, inflation and the interest rate to a monetary policy shock. Since the response is to a unit of the shock, it measures only changes in the transmission of the monetary policy shock over time without taking into account changes in the volatility of the shock, as documented in the previous subsection.

First, the response of the Fed rate to the monetary policy shock is roughly the same throughout the sample period and it is around half a percentage point. The response of output, in contrast, shows a clear trend over time, with responses increasing from around 2.5% to 4% on impact. The response of inflation, on the other hand, displays a U-shape, with inflation being quite responsive to policy in the late 1960s and early 1970s, and in the more recent period. This results are at odds with the findings in Boivin and Giannoni (2006), who find a considerable decrease in the responses of output, inflation and the interest rate to a policy shock in their post-1980 sub-sample, using minimum distance estimator between a structural VAR and a small DSGE model. They attribute this result mainly to the higher estimate of the inflation targeting parameter in their policy rule over the second period.

Figure 15 presents responses to a one standard deviation of the shock and these incorporate the decrease in the volatility of monetary policy shock over the second half of the sample. Instead of increasing responsiveness of output, we now observe somewhat constant response over time with an increase in the end of the sample due to the aggressive policy during the crisis. Interestingly, once one allows for changing size of the shock over time, inflation's response to policy is actually decreasing over time. Furthermore, one standard deviation policy shock results into considerably higher response of the interest rate during Volcker's years than in any other period.

Figures 16 and 17 display the responses to a price mark-up shock. The picture that emerges is that both output and investment become much more responsive to an inflation shock during the Great Moderation period, implying that the same shock to inflation has a relatively more harmful effect on these variables in that period than during the Oil Crises, when inflation was record high. It is also evident from the policy rate response that policy makers retaliated more to a unit of inflation shock after Paul Volcker's appointment as a Federal Reserve chairman.

Figure 14: IRFs to 1 unit monetary policy shock

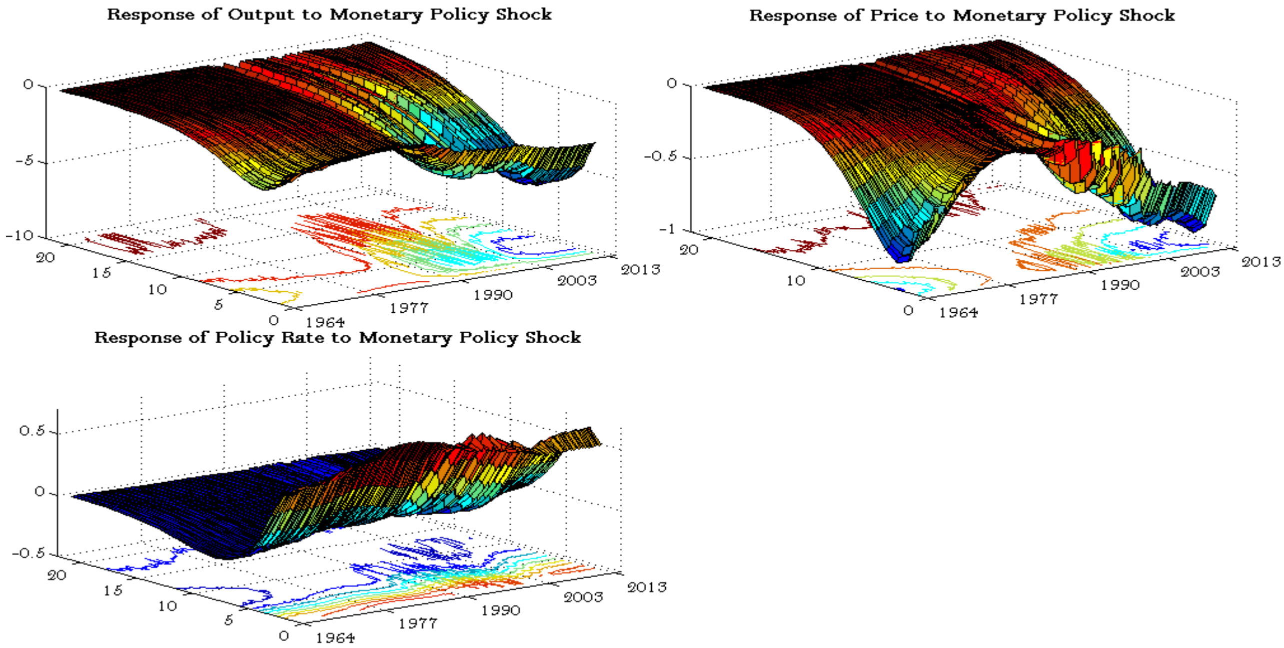


Figure 15: IRFs to 1 st. dev. monetary policy shock

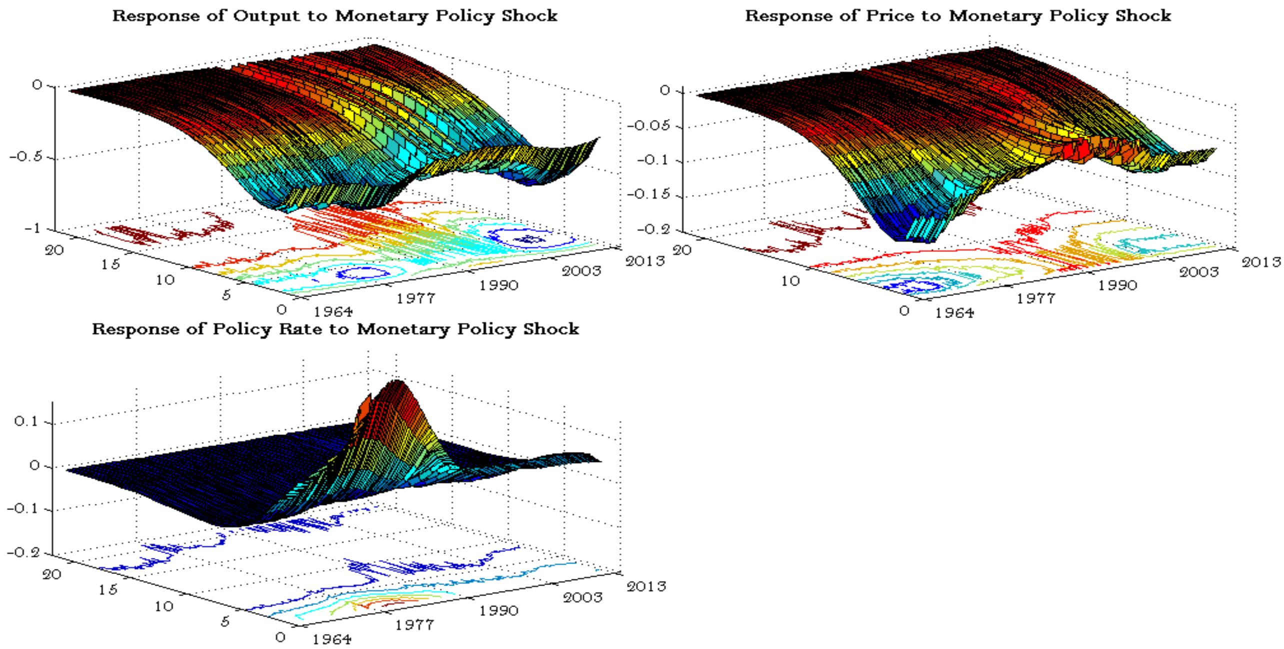


Figure 16: IRFs to 1 unit price mark up shock

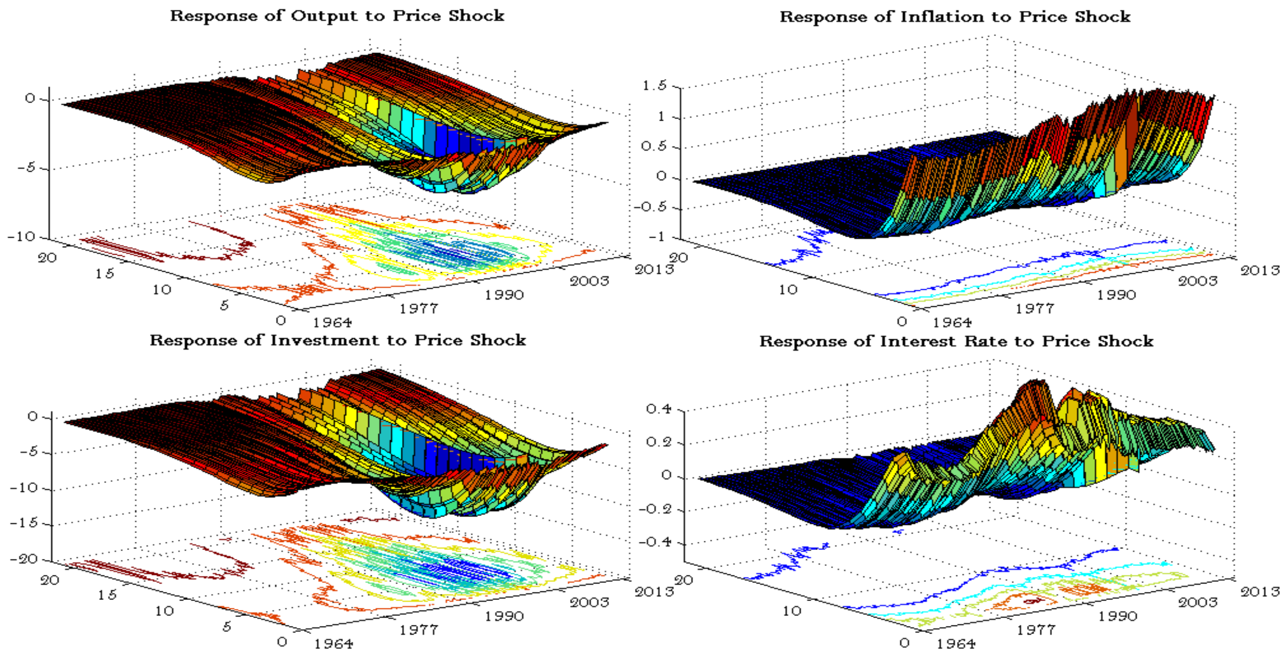


Figure 17: IRFs to 1 st. dev. price mark up shock

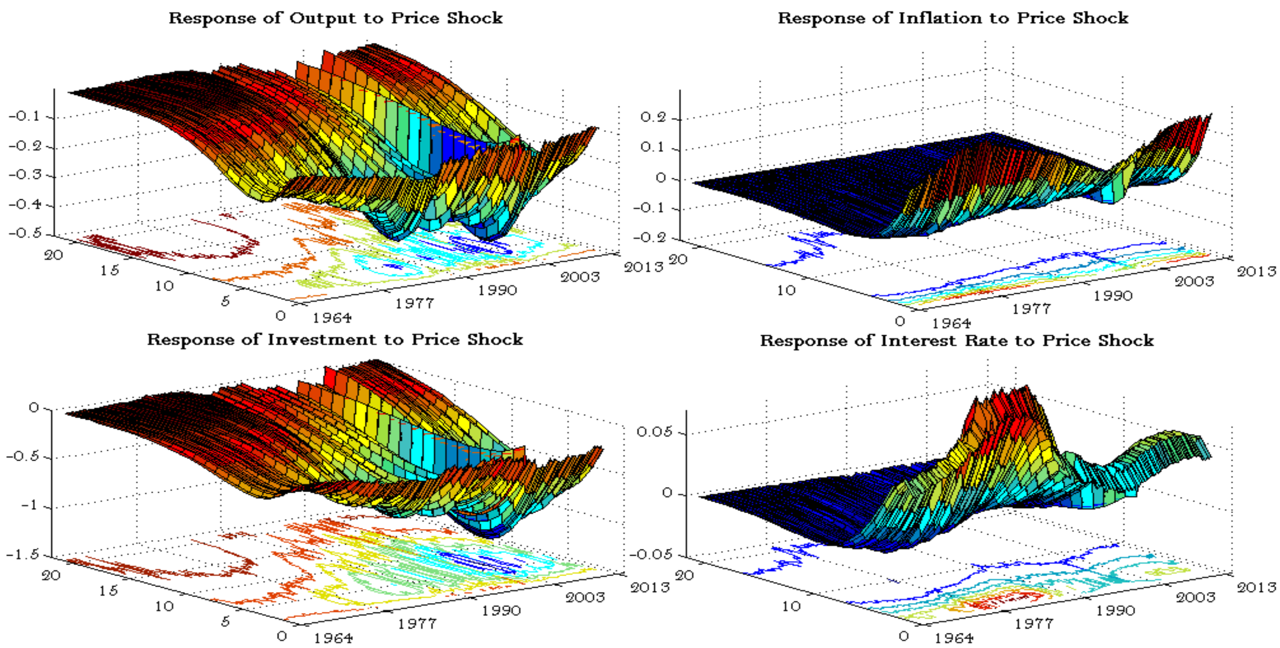


Figure 18: IRFs to 1 unit TFP shock

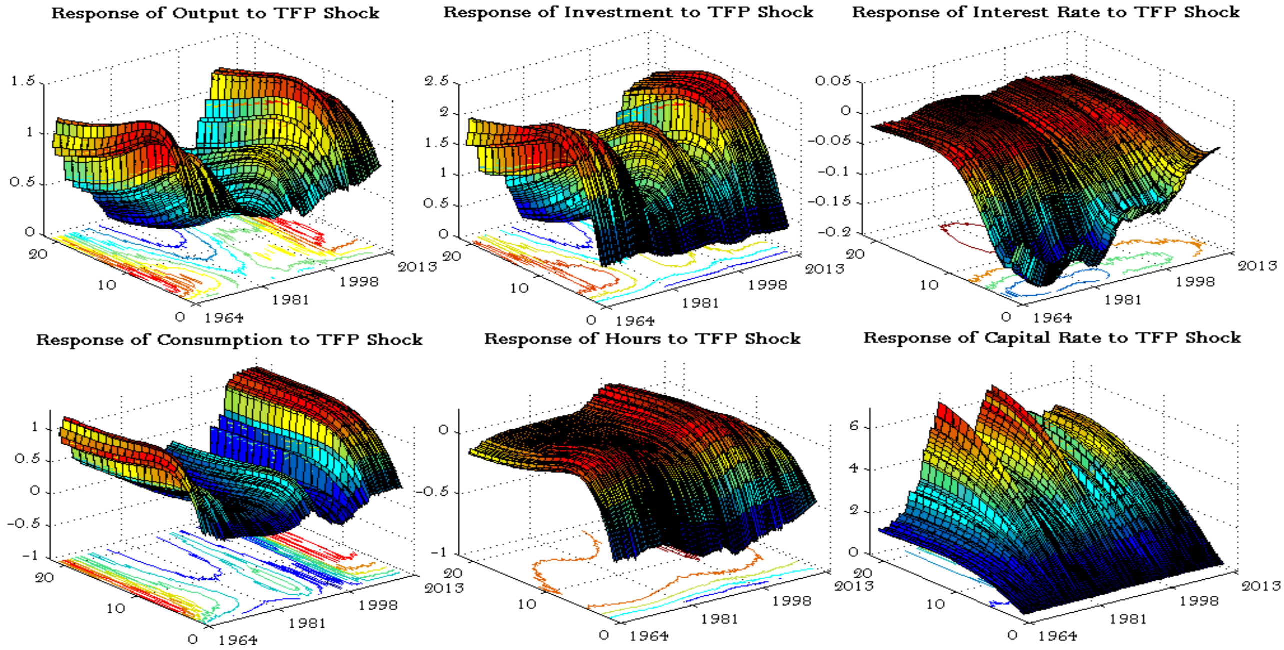
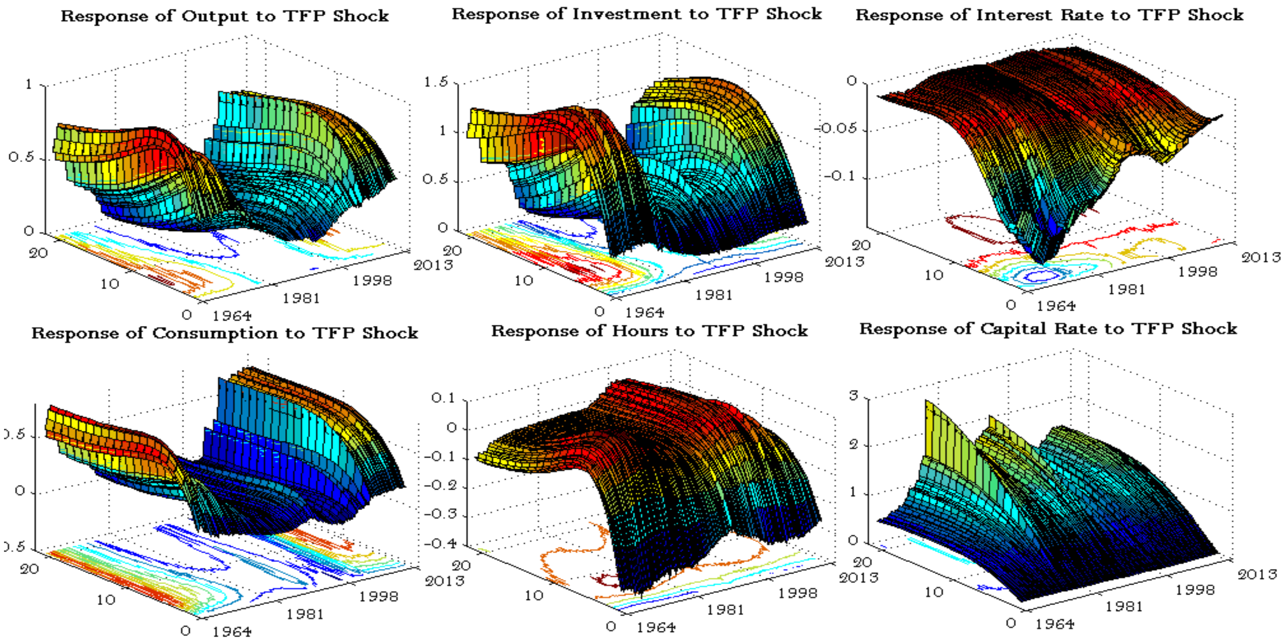


Figure 18 and 19 are the IRFs of selected variables to a unit and a standard deviation of the TFP shock. The most intriguing result that emerges is that the response of all output, consumption and investment, whether one allows for the size of the shock to vary over time or not, is considerably larger in periods characterised by recessions such as the Oil Crisis in early 1970s and the recent crisis, implying an asymmetrically larger effects of TFP shocks in recessions than in booms. The decreasing responsiveness of the interest rate over time could be explained by the Federal Reserve responding less aggressively to output and more aggressively to inflation which is less affected by productivity shocks. The response of hours worked to productivity shock and the resulting implications for the relevance and relative importance of this shock for the business cycle is a much debated topic (Gali (1999)). Once we allow for time variation in the TFP responses, the response of hours remains negative in all periods for all horizons except several periods in the early 1990s when the response changes sign and becomes positive after less than 10 quarters. This could be attributed to the increased persistence of the shock that we uncover during this period. Furthermore, as argued in Smets and Wouters (2007), the habit coefficient is important for explaining the negative effect of TFP on hours and as shown in Figure 11, we obtain low habit persistence in the 1990s which contributes to the weakened duration of the negative effect.Figures 20 and 21 display the responses to a preference shock. It appears that, after allowing for the changing size of the shock over time,

Figure 19: IRFs to 1 st. dev. TFP shock



output, consumption and investment are more responsive to the shock, both on impact and in duration, in periods characterised by recessions, suggesting asymmetric responses to the preference shock in the model.

Finally, Figures 22 and 23 display the responses to a unit and a standard deviation of wage mark up shock respectively. The parameters characterising the labour market during the Great Moderation period, which we discussed in the previous Section, are the reason for the misbehaved impulse response functions during the same period. It is clear that the responses are not smooth over time and the resulting response per unit of the shock of output, inflation and hours becomes essentially zero for all horizons after the beginning of the 1990s. Once we allow for the changing size of the shock, the picture becomes even more distorted, since the standard deviation of the wage mark up shock is itself one of the parameters that are misspecified during the Great Moderation period[28].

[28]The IRFs of selected variables to the remaining shocks (namely, Investment Technology and Government Spending Shocks) can be found in the Appendix 7.2, see Figures 47-50.

Figure 20: IRFs to 1 unit preference shock

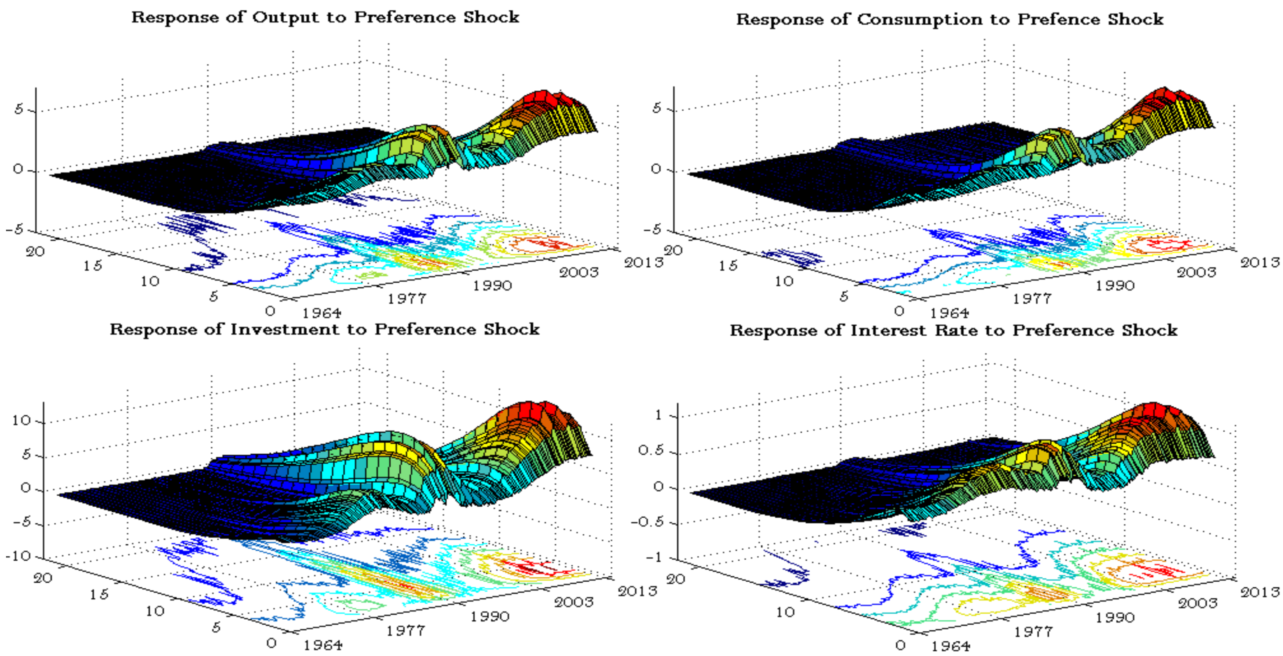


Figure 21: IRFs to 1 st. dev. preference shock

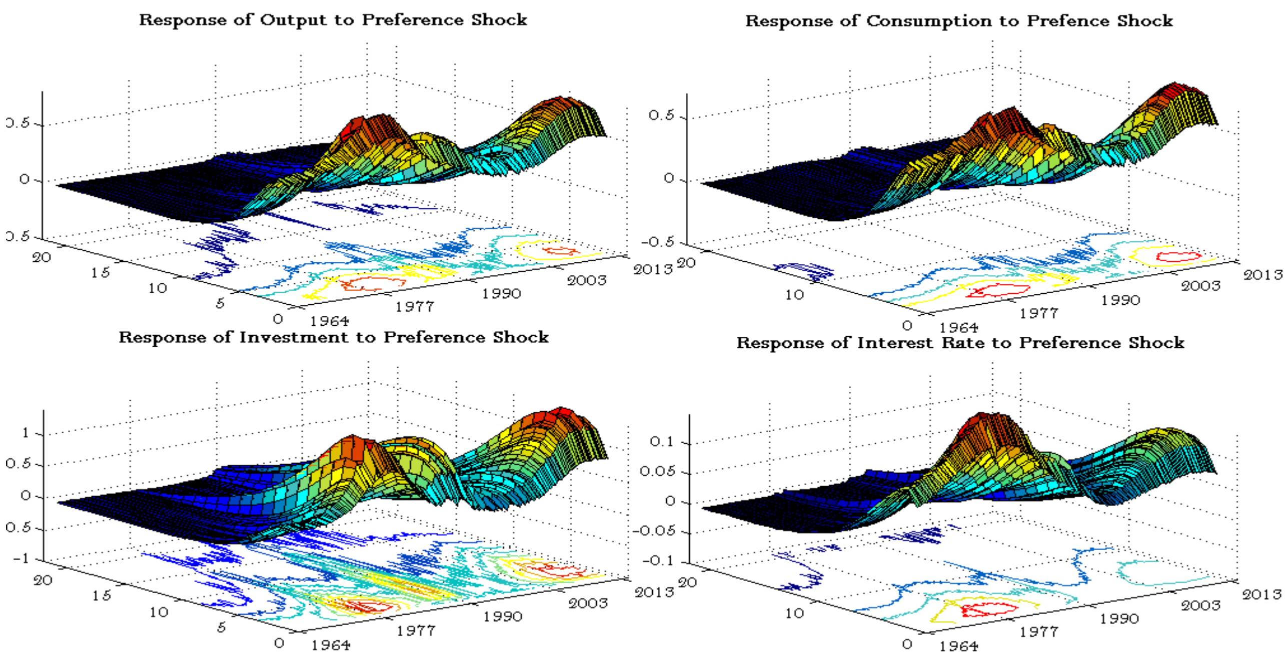
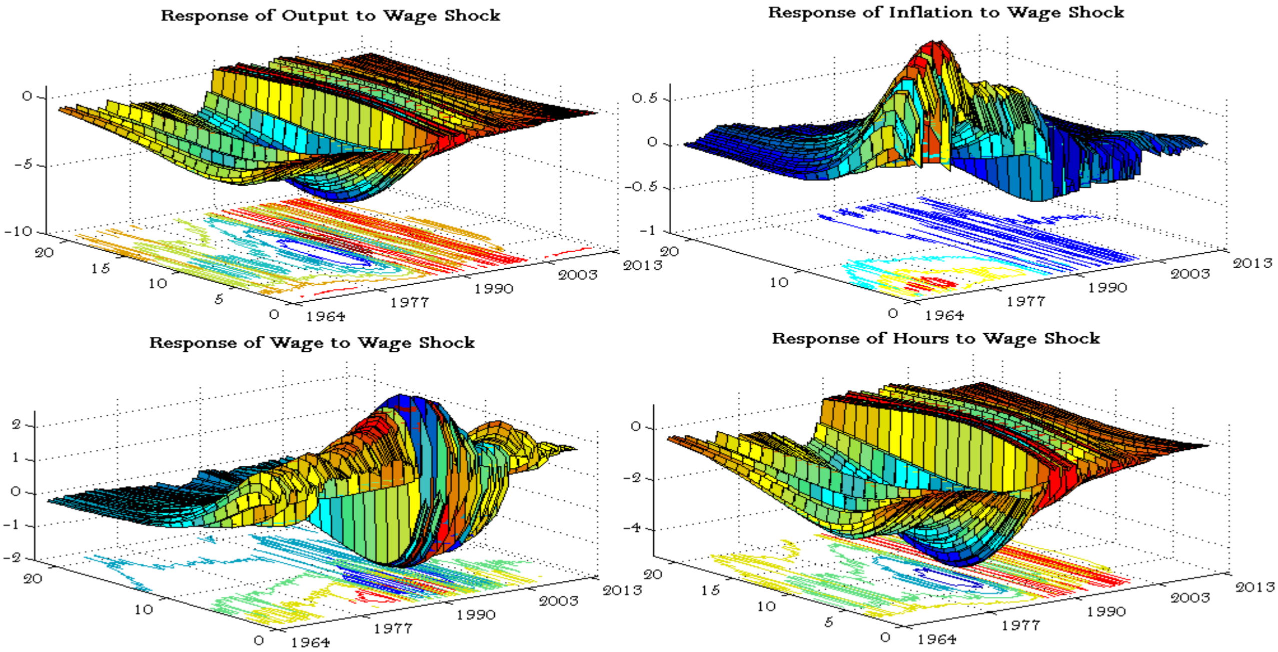


Figure 22: IRFs to 1 unit wage mark up shock



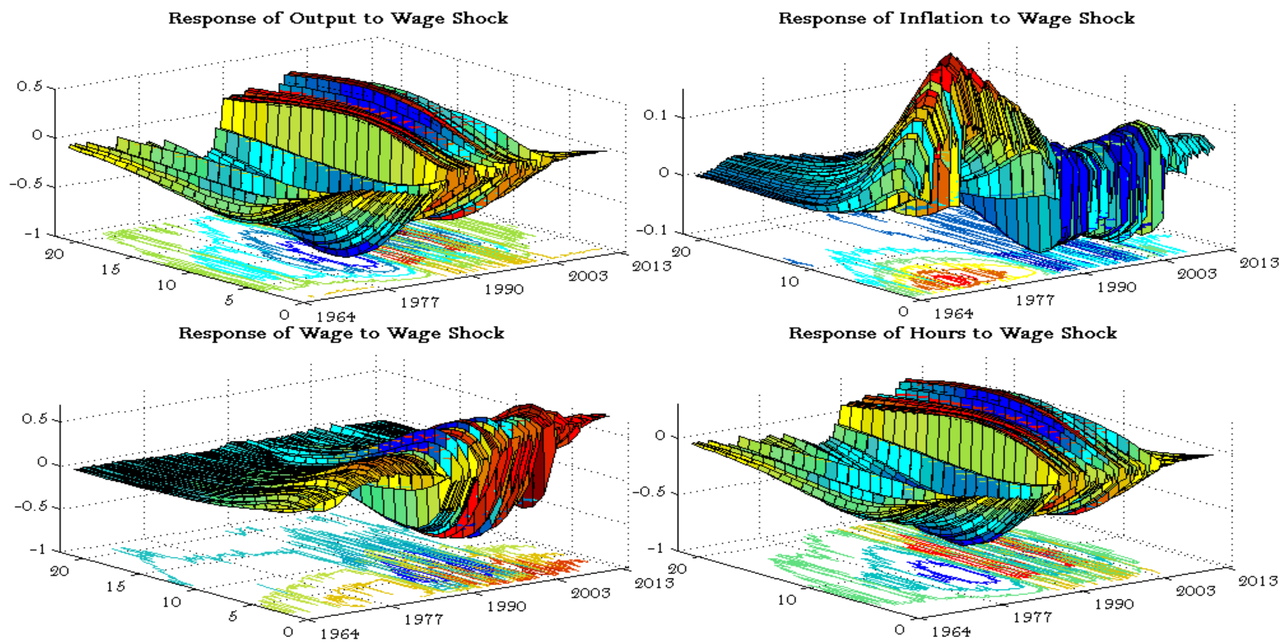
3.7 Forecasting with a time varying DSGE Model

As we discussed in the introduction, the literature has documented evidence of the forecasting accuracy of fixed-parameter DSGE models (Smets and Wouters (2007), Edge et al. (2009), Edge and Guerkaynak (2010), Del Negro and Schorfheide (2013b), Del Negro, Giannoni and Schorfheide (2014)). In this Section we evaluate the relative forecasting performance of our time varying DSGE model. In addition to the fixed-parameter Smets and Wouters (2007) specification, we also compare the forecasting record of the time varying DSGE model against univariate models (AR(1), a Random Walk (RW) and a time varying AR(1)) and multivariate reduced-form models (a BVAR and a time varying stochastic volatility BVAR (TV-SV BVAR)).

The BVAR uses a standard Normal-inverted-Wishart conjugate prior with optimal shrinkage and optimal lag selection as in Carriero, Clark and Marcellino (2015a). The TV-SV BVAR features time varing autoregressive coefficients as in Cogley and Sargent (2002) and stochastic volatility as in Primiceri (2005)[29]. Since it is burdensome to estimate this model for more than three variables and obtain stationary draws from the posterior distribution of the autoregressive coefficients, we limit our TV-SV BVAR to only output growth, inflation and the Fed Funds rate. The TV-AR

[29] The TV-SV BVAR is of lag order one and uses random walk processes for both the autoregressive coefficients and the log volatility. For more details, see for instance Benati and Mumtaz (2007).

Figure 23: IRFs to 1 st. dev. wage mark up shock



model is computed using the non-parametric kernel based method, as in Giraitis et al. (2014)[30].

We employ the algorithm outlined in Section 3.3.2 to generate density forecasts for the observables of the time varying DSGE model.

Since real-time data is limited and only available after 1991[31], we perform the out-of-sample forecasting on final revised data as we would like to be able to assess performance across different periods. Our forecast origins range from 1974Q3 up to 2010Q1 and we compute forecasts for one up to twelve quarters ahead.

We measure accuracy of point forecasts using the root mean squared forecast error (RMSFE) and forecast bias. The accuracy of density forecasts are measured by log predictive scores. We compute the logscore with the help of a nonparametric estimator to smooth the draws from the predictive density obtained for each forecast and horizon. We test whether a model is statistically more accurate than the benchmark with the Diebold and Mariano (1995) statistic computed with Newey-West estimator to obtain standard errors. We provide the results of the Diebold-Mariano test for the RMSFEs and logscores. For the bias, we simply test whether the models' bias is

[30]The model is estimated in each point in time $t : \widehat{\beta}_t = (X'D_tX)^{-1}X'D_tY$ where X contains the lagged dependant variable Y and D_t is a diagonal matrix with the kernel weights of the t^{th} row of the weighting matrix in equation (4) in its main diagonal. The variance of the residuals is also time varying and computed in point t as $\widehat{\sigma}_t^2 = \varepsilon' D_t \varepsilon / tr(D_t)$. Density forecasts are then generated, using wild bootstrap and the last period values $\widehat{\beta}_T$ and $\widehat{\sigma}_T^2$.

[31]Real-time data on compensation is not available prior to 1991.

statistically different from zero.

3.7.1 Point Forecasts

Table 12 presents the absolute performance of the our TV DSGE model (in RMSFEs) and the relative performance of our approach to alternative models over different horizons (numbers smaller than one imply superior performance of the TV DSGE relative to the alternatives). One, two and three stars indicate that we reject the null of equal accuracy in favour of the better performing model at significance levels of 10%, 5% and 1% respectively.

Forecast Performance - RMSFEs: 1974Q3-2011Q3								
	Horizon	Output	Consumption	Investment	Wages	Hours	Inflation	Interest Rate
TV-DSGE	1	0.72	0.66	1.88	0.74	0.63	0.28	0.26
	2	0.79	0.69	2.14	0.74	1.07	0.38	0.40
	4	0.82	0.69	2.30	0.73	1.83	0.48	0.55
	8	0.77	0.67	2.17	0.74	2.69	0.48	0.72
	12	0.77	0.68	2.15	0.78	3.15	0.49	0.81
TV-DSGE/ F-DSGE	1	0.99	0.91**	1.04*	1.01	1.04	1.09***	1.01
	2	0.98	0.87**	1.07*	1.00	1.00	1.14***	1.01
	4	0.98	0.88**	1.08	1.00	0.97	1.17**	0.98
	8	0.97	0.93*	1.05	1.00	0.91	1.07	0.94
	12	0.99	0.96	1.03	0.99	0.89	1.00	0.90
TV-DSGE/ BVAR	1	0.89	0.92	0.99	0.97	1.07*	0.97	0.99
	2	0.94	0.92	0.99	0.99	1.05	0.99	0.97
	4	0.90	0.90	0.98	1.00	0.98	1.02	0.99
	8	0.92	0.90	0.98	0.97	0.87	0.79	0.94
	12	0.93	0.94	0.97	0.98*	0.80	0.72	0.90
TV-DSGE/TV-SV BVAR	1	0.88	-	-	-	-	0.98	1.03
	2	0.97	-	-	-	-	1.00	1.08
	4	0.92	-	-	-	-	1.04	1.05
	8	0.83	-	-	-	-	0.82	0.94
	12	0.81	-	-	-	-	0.74	0.90
TV-DSGE/ AR(1)	1	0.90*	0.94	1.02	1.00	0.88*	1.03	0.96
	2	0.94	0.94	1.01	1.02	0.86*	1.07	0.98
	4	0.97	0.96	1.00	0.99	0.86*	1.08	0.95
	8	0.97	0.97	0.97	0.98	0.82**	0.93	0.84*
	12	0.98	0.99	0.97	0.99	0.80*	0.84	0.80**
TV-DSGE/ RW	1	0.70**	0.72**	0.79*	0.74**	0.52**	0.86*	0.63**
	2	0.71**	0.74**	0.79**	0.72**	0.64**	1.02	0.84
	4	0.69**	0.72**	0.76**	0.69**	0.76**	1.08	0.86
	8	0.70**	0.67**	0.67**	0.74**	0.80**	0.96	0.83*
	12	0.63**	0.66**	0.59**	0.72**	0.80**	0.90	0.80**
TV-DSGE/ TV-AR(1)	1	0.86**	0.87**	0.93	0.97	0.86**	1.01	0.94
	2	0.89*	0.84*	0.86*	0.98	0.82**	1.04	0.98
	4	0.90*	0.82	0.81*	0.96	0.80**	1.07	0.86
	8	0.88**	0.74	0.71**	0.95**	0.73**	0.94	0.73**
	12	0.87**	0.73	0.66*	0.94**	0.68**	0.89	0.64**

Table 12: RMSFEs. The figures under TV-DSGE model are absolute RMSFEs, the figures under the remaining models are ratios of RMSFEs of TV-DSGE over the alternatives. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

There are some gains from using the time varying model over the standard fixed parameter DSGE for most variables but the differences are small and rarely significant. One exception is inflation: the time varying model performs significantly worse than the fixed-parameter specification.

The model also outperforms the standard BVAR, which confirms previous findings (e.g. Smets and Wouters (2007), Adolfson, Andersson, Linde, Villani and Vredin (2007), Christoffel, Coenen and Warne (2010)). Moreover, we find superior performance for output growth over the TV-SV BVAR. Finally, the TV DSGE model strongly outperforms the univariate models.

In order to better understand the strengths and weaknesses of our approach, we further investigate the point forecast accuracy by splitting our sample of forecasts into subsamples corresponding roughly to three distinct periods in U.S. recent economic history: namely, the Oil Crisis or Great Inflation period (at least the end of it, ranging 1974Q2:1982Q4), the Great Moderation (1983Q1:2005Q4) and the recent financial crisis (2006Q1:2011Q3). Table 13 presents the relative RMSFE performance of our approach and Table 14 displays the forecast bias for per capita GDP growth, inflation and the interest rate during the three periods. For the Oil Crisis period, our method is superior to the two BVARs and comparative to the standard DSGE approach for output. When it comes to forecast bias in the Great Inflation Period, both DSGE models strongly and significantly underestimate inflation, but in relative terms our model does poorly, resulting in significantly worse RMSFE performance. The reason for this result is the relatively small sample size at this point and hence, little advantage in down-weighting past data. Interestingly, the two BVARs deliver unbiased inflation forecasts, but systematically underestimate output growth.

Forecast Performance - RMSFEs										
		Oil Crisis			Great Moderation			Recent Crisis		
		1974Q3-1982Q4			1983Q3-2005Q4			2006Q1-2011Q3		
	Horizon	Output	Inflation	Int Rate	Output	Inflation	Int Rate	Output	Inflation	Int Rate
TV-DSGE	1	0.96	0.46	0.46	0.65	0.18	0.16	0.60	0.26	0.10
	2	1.16	0.67	0.71	0.62	0.20	0.27	0.72	0.29	0.18
	4	1.22	0.90	0.91	0.64	0.22	0.39	0.70	0.35	0.33
	8	1.13	0.89	1.17	0.61	0.25	0.53	0.69	0.28	0.48
	12	1.18	0.91	1.34	0.57	0.27	0.56	0.71	0.19	0.53
TV-DSGE/ F-DSGE	1	0.98	1.15***	1.03	1.02	0.97	0.99	0.90	1.13*	0.76**
	2	1.08**	1.19***	1.05	0.92**	0.96	0.96	0.86	1.21	0.79*
	4	1.06*	1.19*	1.01	0.98	0.98	0.96	0.79	1.34	0.79
	8	1.01	1.11	0.99	1.00	0.98	0.99	0.82	0.91	0.63
	12	1.03	1.05	0.97	0.97	0.96	1.04	0.88	0.54	0.52
TV-DSGE/ BVAR	1	0.81	1.07	1.04	1.04	0.76**	1.03	0.76	1.13	0.47*
	2	1.04	1.12	0.98	0.95	0.69**	1.13*	0.72	1.14	0.53
	4	0.97	1.26	0.99	1.01	0.57**	1.16*	0.62*	1.05	0.65
	8	0.96	1.00	0.96	1.09	0.50**	1.03	0.64	0.62	0.66
	12	0.91	0.95	1.00	1.09	0.46**	0.92	0.77	0.37	0.54
TV-DSGE/ TV-SV BVAR	1	0.83	1.03	1.02	0.97	0.84*	1.06	0.80	1.09	0.86*
	2	1.12	1.10	1.10	0.89	0.73*	1.06	0.81	1.22	0.82
	4	1.02	1.21	1.09	0.88	0.61	1.02	0.76	1.32	0.83
	8	0.81	0.88	0.95	0.91	0.62	1.01	0.75	1.02	0.73
	12	0.76	0.77	0.94	0.93	0.62	0.91	0.77	0.78	0.67

Table 13: RMSFEs. The figures under TV-DSGE model are absolute RMSFEs, the figures under the remaining models are ratios of RMSFEs of TV-DSGE model over the alternatives. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

Although, as argued earlier, our method does on average worse for inflation over the full forecast sample, during the Great Moderation, we obtain better performance. The Great Moderation was characterised by low and stable inflation and very low business cycle volatility. It is clear from the forecast bias, that both BVAR models and the standard DSGE model significantly overestimate inflation during this period. This is the case, since the standard DSGE and the fixed coefficient BVAR models, in order to generate projections, use samples which contain the entire Oil Crisis period, characterised by very high inflation. Our method, on the other hand, obtains better performance and remains virtually unbiased because of its way of down-weighting distant data. The TV-SV BVAR features time variation in the autoregressive coefficients of the VAR, so it is surprising that is fails to capture the structural change in inflation dynamics during the Great Moderation and also systematically overestimates inflation. This could be due to the way time variation enters the model; our approach models time variation non-parametrically, and it is more robust to misspecifications in the stochastic processes for the time varying parameters that the TV-SV BVAR utilises (for further discussion and Monte Carlo evidence, see Section 2.9 and Appendix 7.1.11). All models deliver similar, in terms of RMSFEs, output forecasts during the Great Moderation; however, both DSGE models underestimate output.

Forecast Performance - Bias										
		Oil Crisis			Great Moderation			Recent Crisis		
		1974Q3-1982Q4			1983Q3-2005Q4			2006Q1-2011Q3		
	Horizon	Output	Inflation	Int Rate	Output	Inflation	Int Rate	Output	Inflation	Int Rate
TV-DSGE	1	0.11	-0.23**	-0.06	-0.10	0.00	-0.01	-0.04	-0.08	0.01
	2	0.15	-0.39**	-0.14	-0.19**	0.01	0.00	0.07	-0.13*	0.03
	4	0.18	-0.60**	-0.34	-0.25**	0.03	0.02	0.16	-0.17	0.09
	8	-0.02	-0.63	-0.58	-0.25**	0.02	0.03	0.18	-0.12	0.24
	12	-0.06	-0.66	-0.88	-0.18	0.01	0.05	0.13	-0.04	0.45
F-DSGE	1	0.06	-0.18**	-0.07	-0.18**	0.02	-0.01	0.31**	-0.01	0.04
	2	0.04	-0.32**	-0.17	-0.30**	0.04	-0.01	0.48**	0.00	0.11*
	4	0.09	-0.52**	-0.41	-0.34**	0.08**	-0.01	0.55*	0.06	0.28*
	8	-0.02	-0.58*	-0.69	-0.27**	0.11*	0.03	0.49	0.19	0.64
	12	-0.04	-0.63	-1.00	-0.20	0.13*	0.10	0.39	0.29	0.96
BVAR	1	-0.41**	0.03	-0.06	0.02	0.11***	-0.03	0.33**	0.02	-0.11**
	2	-0.47**	0.05	-0.18	0.03	0.17***	-0.01	0.61**	0.03	-0.13
	4	-0.28	0.05	-0.44*	0.01	0.27***	0.06	0.86**	0.05	-0.04
	8	-0.16	0.13	-0.62	-0.07	0.42***	0.22	0.84*	0.21	0.35
	12	-0.13	0.15	-0.78	-0.09	0.51**	0.37*	0.59	0.38	0.79
TV-SV BVAR	1	-0.25	-0.05	-0.01	-0.05	0.07***	0.02	0.40***	-0.01	0.00
	2	-0.36*	-0.04	-0.03	-0.06	0.12***	0.04	0.55**	0.00	0.03
	4	-0.42*	0.01	-0.07	-0.06	0.17**	0.09	0.61*	0.01	0.11
	8	-0.67*	0.22	-0.03	-0.02	0.19*	0.15	0.61	0.07	0.34
	12	-0.73	0.33	-0.14	0.04	0.20	0.22	0.56	0.14	0.60

Table 14: Forecast Bias. The table reports forecast bias, computed as the mean forecast error. '*', '**' and '***' indicate rejection of the null of zero bias against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

The period of the recent financial crisis (2007-2009) has been a subject to many discussions.

This crisis generated serious critiques for the forecasting literature, for instance, Wieland and Wolters (2011), Del Negro and Schorfheide (2013b) provide evidence that DSGE models not only failed to predict it, but even once the crisis had started, failed to forecast the trough and quickly returned the economy to positive growth rates. During the 2006-2011 period, which overlaps with the recent crisis period and subsequent recovery, our model outperforms the standard DSGE model and both BVARs for all horizons and even with a small sample size of 23 observations, manages to deliver some statistically significant improvements. This could also be seen from the forecast bias where all alternative models considerably overestimate output, but our approach impressively delivers unbiased output forecasts at one step ahead, compared to a bias of around 0.31% quarterly GDP growth for the fixed parameter DSGE model, 0.33% for the BVAR and 0.40% for the TV-SV BVAR. Our interpretation of this result is similar to before; both the BVAR model and the standard DSGE model use as in-sample period data from Oil Crisis and the Great Moderation in order to generate forecasts for the recent crisis, while our method also makes use of these data but discounts it and gives more weight to recent observations. The resulting trend coefficient, $\overline{\gamma}$, as seen in Section 3.5, falls considerably after 2007. This parameter is important and can have substantial effect on the model's forecasts, as it enters as an intercept in the measurement equation for output, consumption, investment and wage growth. For inflation, while RMSFE performance during the same period is relatively similar to the standard DSGE model (worse in the short run and better at longer horizons), the forecast bias is negative for the TV DSGE model while positive for the standard DSGE. Another interesting result is that, our method, while delivering similar interest rates forecasts during the previous subsamples, delivers very large and statistically significant improvements over all models during the crisis. Our TV DSGE model contains a Taylor rule with changing parameters and in particular, during the crisis period with interest rates close to the Zero Lower Bound (ZLB), our interest rate smoothing coefficient jumps to levels near 0.9. This delivers interest rate forecasts that are close to a random walk model while inflation targeting and output gap values have a reduced effect.

To summarise, in periods, in which serious structural change is present, such as the recent one, the forecast errors of the TV DSGE model are relatively smaller and the model is more robust to forecast bias resulting from the presence of the structural change in the in-sample period.

3.7.2 Density Forecasts

Table 15 accesses the quality of the density forecasts measured by logscores of the predictive density. The table displays absolute log predictive score for the TV DSGE model and differences in logscores over the alternative models, so numbers greater than zero imply superior performance of our approach.

A few comments are in order. First, it is clear that overall our method outperforms considerably and statistically significantly the standard DSGE for most variables and horizons. Interestingly, our simple univariate TV AR(1) also delivers density performance superior to that of the standard DSGE model. These results are important as they imply that, while allowing for parameter drift results in moderate gains for point accuracy and only for some variables and periods, it results in significantly improved density forecasts. One way to look at this is to infer about the importance of stochastic volatility. As seen in Section 3.5, the uncovered time variation in the standard deviations of the shocks in the structural model is substantial and previous studies have confirmed this result (Primiceri (2005), Sims and Zha (2006), Justiniano and Primiceri (2008)). Since volatility is inherently time varying and subjected to structural change, it is clear that conditioning on the most recent values of the volatility of the shocks when simulating the density of forecasts, as outlined in Section 3.3.2, will deliver better forecast uncertainty. Since the TV-SV BVAR also allows for changing volatility, it is unsurprising that it delivers very similar density forecast performance.

	Horizon	Output	Consumption	Investment	Wages	Hours	Inflation	Interest Rate
Forecast Performance - Log Predictive Score: 1974Q3-2011Q3								
	Horizon	Output	Consumption	Investment	Wages	Hours	Inflation	Interest Rate
TV-DSGE	1	-1.14	-1.05	-2.00	-1.28	-0.97	-0.02	0.30
	2	-1.20	-1.06	-2.17	-1.10	-1.50	-0.25	-0.30
	4	-1.23	-1.10	-2.28	-1.09	-2.15	-0.45	-0.82
	8	-1.22	-1.07	-2.22	-1.11	-2.65	-0.52	-1.19
	12	-1.21	-1.05	-2.23	-1.19	-2.81	-0.53	-1.36
TV-DSGE-F-DSGE	1	0.05**	0.05	0.03	0.00	-0.04	0.02	0.34***
	2	0.06**	0.12***	0.02	0.12	0.03	-0.01	0.25**
	4	0.06**	0.11***	-0.03	0.05	0.06	0.00	0.13
	8	0.07*	0.06	-0.04	0.05	0.19	0.07	0.08
	12	0.07	0.07*	0.00	0.03	0.42	0.15**	0.18
TV-DSGE-BVAR	1	0.09	0.03	0.06	-0.02	-0.10**	0.14**	0.39***
	2	0.08	0.05	0.10	0.15	-0.01	0.16*	0.22**
	4	0.11	0.07	0.09	0.11	0.02	0.20*	0.01
	8	0.05	0.06	0.03	0.15	0.13	0.34**	0.01
	12	0.04	0.08*	0.07	0.12	0.42	0.45**	0.09
TV-DSGE-TV-SV BVAR	1	0.04	-	-	-	-	-0.01	-0.05
	2	0.03	-	-	-	-	-0.06	-0.06
	4	0.00	-	-	-	-	-0.09	-0.07
	8	0.09	-	-	-	-	0.01	0.05
	12	0.10	-	-	-	-	0.08	0.06
TV-DSGE-AR(1)	1	0.03	-0.05	0.09*	0.05	0.16**	0.12*	0.28***
	2	-0.04	-0.09	-0.04	0.19***	0.25*	0.11*	0.19**
	4	-0.05	-0.12	-0.06	0.24***	0.30	0.18**	0.10
	8	-0.09	-0.14**	-0.05	0.26***	0.43*	0.29**	0.24
	12	-0.11	-0.13*	-0.05	0.23**	0.48**	0.40**	0.37
TV-DSGE-RW	1	0.25***	0.23**	0.35***	0.39***	1.02***	0.34***	0.79***
	2	0.33***	0.26**	0.21	0.37***	0.69***	0.13*	0.41***
	4	0.51***	0.49***	0.26	0.39***	0.41**	0.17*	0.23
	8	0.84***	0.84***	0.57***	0.55***	0.44**	0.38***	0.15
	12	1.06***	1.07***	0.74***	0.66***	0.29	0.57***	0.13
TV-DSGE TV-AR(1)	1	0.12*	0.11	0.20**	0.12	0.23***	0.13**	0.29***
	2	0.10	0.06	0.12	0.27***	0.42***	0.10	0.34**
	4	0.09	0.06	0.20	0.25***	0.70**	0.14	0.40**
	8	0.08	0.05	0.34**	0.26***	1.42***	0.16	0.82**
	12	0.04	0.00	0.34**	0.21**	1.93***	0.16	1.12**

Table 15: Log Predictive Score. The figures under TV-DSGE model are absolute log predictive scores, the figures under the remaining models are differences of RMSFEs over the TV-DSGE model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

Table 16 further investigates the relative density performance of our approach over the sub-sample periods. It is clear that our method delivers better forecast uncertainty than the standard DSGE and the standard BVAR over the Great Moderation period, which is unsurprising. The in-sample that the two fixed coefficient models contain is the high volatility period of the Oil Crises, hence, as anticipated, density forecasts during the Great Moderation generated with these in-samples are worse. On the other hand, our TV DSGE model as well as the TV-SV BVAR account for this reduced uncertainty (our approach - by kernel down-weighting of Oil Crises data and the TV-SV BVAR - by fitting random walk state equations for the volatility paths) and hence deliver similar and superior performance over the fixed parameter models.

Forecast Performance - Log Predictive Score										
		Oil Crisis			Great Moderation			Recent Crisis		
		1974Q3-1982Q4			1983Q3-2005Q4			2006Q1-2011Q3		
	Horizon	Output	Inflation	Int Rate	Output	Inflation	Int Rate	Output	Inflation	Int Rate
TV-DSGE	1	-1.44	-0.71	-0.77	-1.04	0.27	0.55	-1.07	-0.09	0.84
	2	-1.57	-1.13	-1.31	-1.05	0.06	-0.07	-1.25	-0.20	0.22
	4	-1.65	-1.44	-1.55	-1.09	-0.10	-0.65	-1.21	-0.37	-0.46
	8	-1.58	-1.38	-1.98	-1.10	-0.26	-0.99	-1.15	-0.32	-0.86
	12	-1.60	-1.38	-2.02	-1.10	-0.29	-1.23	-1.11	-0.26	-0.94
TV-DSGE-F-DSGE	1	-0.04	-0.17**	0.21	0.09***	0.13**	0.35***	0.02	-0.12	0.48***
	2	-0.08**	-0.21*	0.42	0.13***	0.09**	0.17	0.02	-0.12	0.28*
	4	-0.09**	-0.16	0.59	0.10***	0.11**	-0.04	0.13*	-0.18	0.10
	8	-0.06*	-0.07	0.31	0.10**	0.14**	-0.12	0.13	0.02	0.45
	12	-0.05*	-0.02	0.85	0.10	0.19**	-0.32	0.11	0.21	1.02
TV-DSGE-BVAR	1	0.42	-0.01	0.34	-0.05	0.27***	0.33***	0.12	-0.10	0.66**
	2	-0.01	0.02	0.22	0.00	0.28***	0.14	0.49	-0.08	0.48
	4	0.06	-0.13	0.04	-0.02	0.39***	-0.07	0.66	-0.02	0.21
	8	0.00	-0.10	0.04	-0.03	0.50***	-0.07	0.38	0.34*	0.26
	12	0.08	-0.05	0.40	-0.03	0.60**	-0.18	0.23	0.60	0.61
TV-DSGE-TV-SV BVAR	1	0.19	-0.07	-0.24	-0.05	0.02	-0.03	0.18	-0.02	0.10
	2	-0.01	-0.20	-0.25	-0.07	0.01	-0.07	0.42	-0.11	0.23
	4	-0.05	-0.38	-0.16	-0.10	0.05	-0.16	0.38	-0.17	0.37
	8	0.10	-0.05	-0.08	-0.09	0.09	-0.08	0.70	-0.18	0.67
	12	0.32	0.17	0.06	-0.05	0.09	-0.14	0.33	-0.07	0.76

Table 16: Log Predictive Score. The figures under TV-DSGE model are absolute log predictive scores, the figures under the remaining models are differences of RMSFEs over the TV-DSGE model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

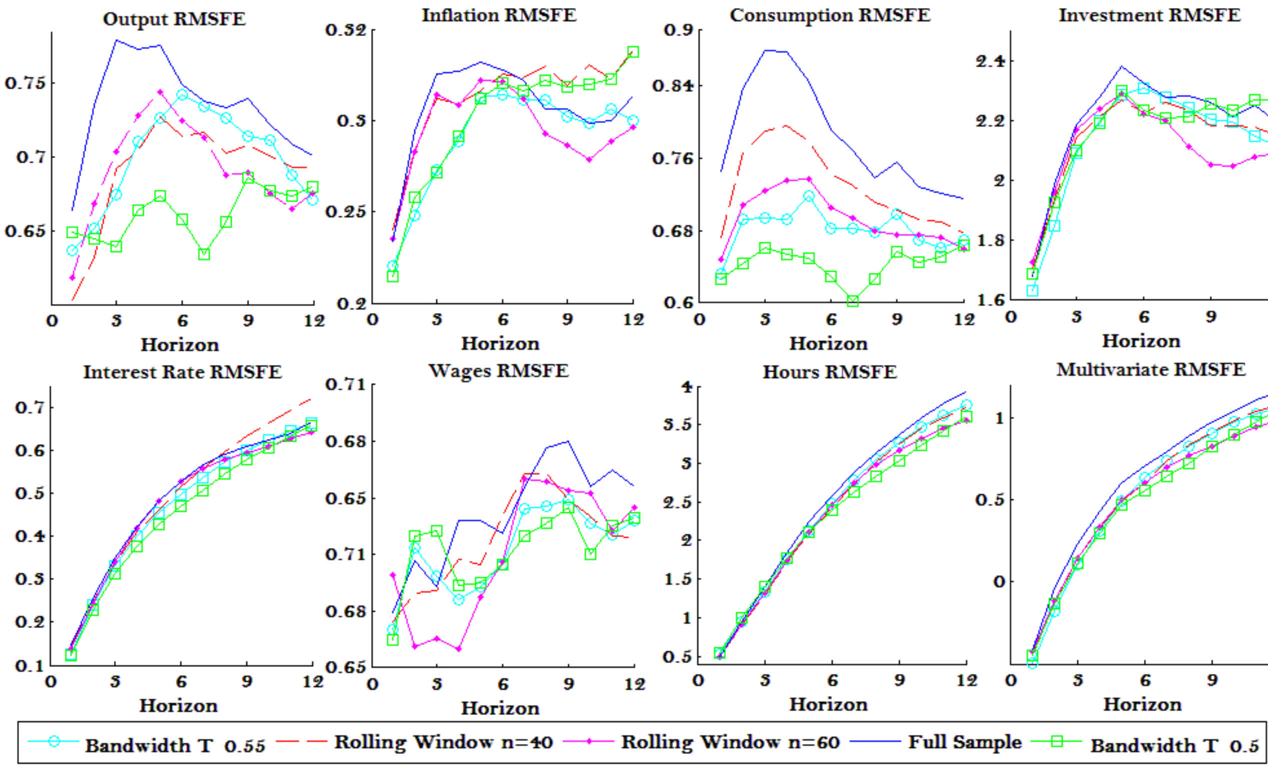
3.7.3 Robustness Checks

In this Section, we investigate the impact of some of our assumptions on the forecasting performance of the time varying DSGE model. We exploit the impact of different bandwidth sizes and the use of the rolling windows method. Figure 25 plots the RMSFEs[32] for the fixed-parameter DSGE model and the time varying DSGE estimated under different assumptions. We include our baseline case with the normal kernel method and bandwidth equals to $T^{0.5}$, and also the case the bandwidth of $T^{0.55}$. We also consider flat kernels which are equivalent to rolling windows of 40 and 60 observations. The results in Figure 25 support our baseline estimation method since they imply a forecasting performance that is superior to alternative specifications.

Robustness Check. Comparison of RMSFEs obtained with bandwidths $H = T^{0.5}, T^{0.55}$ and rolling windows of size 40 and 60.

[32]For computational time considerations, the robustness checks have only been computed at the mode of the parameter posterior. Furthermore, the sample size of the forecasts is smaller than in the comparison in the previous Section due to use of larger in-sample periods by the wider rolling windows.

Figure 24: Robustness Check



3.8 Summary

This Chapter develops a quasi-Bayesian local likelihood method to accommodate time variation in DSGE models' parameters, appropriately weighting the sum of log likelihoods of the sample with weights generated by a kernel function. The method can be applied to any DSGE model that has a state-space representation. The empirical application presented uncovers some time variation in the Smets and Wouters (2007) model's parameters and points to potential misspecification in the labour market of the model during to the low volatility environment of the Great Moderation period. When it comes to forecasting, our estimation procedure is no more computationally intensive than estimating a DSGE model with fixed parameters and, as demonstrated in the forecasting exercise, can deliver some gains in the forecast performance both for point and density projections especially in the presence of serious structural change such as the recent financial crisis.

4 Time Varying DSGE model with Financial Frictions

4.1 Introduction

The previous Chapter discussed the success of dynamic stochastic general equilibrium (DSGE) for policy analysis and macroeconomic forecasting. In particular, following Smets and Wouters (2007), many authors evaluated the DSGE models's forecasting performance, providing evidence that they can produce accurate forecasts of output growth and inflation in real time (Edge and Guerkaynak (2010), Woulters (2012) and Del Negro and Schorfheide (2013b)). However, the recent financial crisis has posed a serious challenge to macroeconomic modelling and forecasting. Perhaps the most important aspect of this challenge is the inability of standard DSGE models to accommodate the impact of developments in the financial sector on the rest of the economy. Based on the seminal work of Bernanke, Gertler and Gilchrist (1999), various authors have exploited financial channels in a DSGE structure as a way of improving the fit of the DSGE model to the 2008-2009 global financial crisis, including Christiano, Motto and Rostagno (2014), Del Negro and Schorfheide (2013b) and Del Negro, Hasegawa and Schorfheide (2014). Interestingly, Del Negro, Hasegawa and Schorfheide (2014) find that the Smets and Wouters (2007) model with financial frictions, while delivering relatively better forecasts during the crisis, performs worse in tranquil periods than the model without financial frictions. This is consistent with evidence of the changing predictive power of various economic and financial indicators on U.S. output and inflation (Stock and Watson (2003)). Even if asset prices are, on average, poor indicators of economic activity, their predictive power should have increased during the recent financial crisis. For example, Gilchrist and Zakrajsek (2011) and Philippon (2009) argue that the predictive power of corporate bond credit spread for the business cycle and economic activity reveal the potential of bond markets to signal (even more accurately than stock markets) the decline in fundamentals prior to the 2007-2008 business cycle downturn.

Incorporating a financial channel in a DSGE model may not be enough to address the effect that structural changes in the underlying economy might have on preference parameters and on exogenous shock processes. As argued in Chapter 3, a standard assumption in the literature has been that the DSGE parameters are structural in the Lucas sense, that is, they are invariant to both policy and structural shocks. However, long term cultural or technological shifts might result in slow parameter variation. While DSGE analysis focuses primarily on business cycle frequency, parameter drift is potentially of great importance when considering sample periods of over 40 years, which are routinely used for estimation and calibration of DSGE models. A related issue is the extent to which all parameters of medium-sized DSGE models are equally immune to the Lucas critique. While parameters such as households' discount factor with distinct microfoundations may be unaffected to long run change, other parameters associated with rigidity dynamics have a reduced-form flavour and may be more vulnerable to technological or social change, or other factors. Even if one believes in the structural nature of DSGE parameters, it is important that one recognises at least the possibility of time variation in the parameters when estimated over long time periods.

In this Chapter, we employ the approach developed in Chapter 3 to investigate the changing nature of the effect of financial frictions to the rest of the economy. The model we investigate is a Smets and Wouters (2007) model with an added financial sector as in Bernanke et al. (1999) and Del Negro and Schorfheide (2013b). The advantage of the specification discussed in this Chapter is that the importance of the financial frictions for macroeconomic variables depends on a preference parameter and on the stochastic properties of the new financial friction shock. By looking at the possibility of time variation in these parameters, while also allowing all other DSGE parameters to change over time, we can measure whether the significance of financial frictions change over time. We find that the parameter that triggers the transmission of financial frictions to the economy remains relatively constant during the entire sample period we analyse. However, the volatility of the financial friction shock rises dramatically during the 2007-2011 period. This new finding contributes to the debate between 'Good Luck' versus 'Good Policy' when explaining the Great Moderation (Gali and Gambetti (2009), Benati and Surico (2009), Sims and Zha (2006)). We provide evidence that the financial frictions shock was muted during the 1985-2007 period. Note that our model presents arguments in favour of changes in the volatility of the shocks while also allowing for changes in the parameters of the policy rule. As a consequence, this Chapter produces a new source of evidence of 'Good Luck' during the Great Moderation period while also allowing for a 'Good Policy' channel. Related investigation of changes in the volatility of financial shocks over time is presented in Fuentes-Albero (2014), where a Smets and Wouters (2007) DSGE model with financial frictions is estimated with constant parameters over different subsamples and the breaks in the volatility of the residuals of the model are dated. The author finds that the size of the financial shocks has increased over time, while their importance in explaining non-financial variables has remained relatively unchanged. Cardani, Paccagnini and Villa (2015) estimate a Smets and Wouters (2007) model with banking intermediation and report the posterior means of the parameters over time implied by their recursive forecasting scheme, providing further evidence of the changing importance of the financial shock in their model.

This Chapter also exploits the forecasting performance of the time-varying DSGE model with financial frictions, extending the results of Del Negro and Schorfheide (2013b) and Kolasa and Rubaszek (2015), who use only fixed parameters specifications.

The Chapter is organised as follows. Section 4.2 describes the DSGE model used in the empirical applications in Section 4.3. Finally, Section 4.4 concludes and Appendix 7.3 contains data description, priors, additional results and robustness checks.

4.2 The DSGE model with financial frictions

The DSGE model with financial frictions combines the Smets and Wouters (2007) model (SW), which extends a small-scale monetary RBC model with sticky prices (such as Goodfriend and King (1997), Rotemberg and Woodford (1997), Woodford (2003), Ireland (2004) and Christiano et al. (2005)), with financial frictions as in Bernanke et al. (1999). In addition to the sticky prices, the SW model also includes additional shocks and frictions, featuring sticky nominal price and wage settings with backward inflation indexation, investment adjustment costs, fixed costs in production, habit formation in consumption and capital utilization. Our complete log-linearised specification of the model is described in Appendix 7.3. It differs from the financial friction specification in Kolasa and Rubaszek (2015) and Del Negro and Schorfheide (2013b) in that we are using a deterministic rather than stochastic trend in productivity.

In comparison with the SW model, the main difference of the model discussed in this Chapter is the inclusion of a financial sector from where entrepreneurs borrow funds to finance their projects. To prevent entrepreneurs to accumulate enough for self-financing, the model assumes that a constant proportion of them dies each period. The success of the entrepreneurs' projects depend on both aggregate and idiosyncratic shocks. While entrepreneurs observe the impact of both types of shocks, the banks do not observe idiosyncratic shocks. The financial intermediary faces a standard agency problem in writing the optimal contract to lend to the entrepreneurs. The bank charges a finance premium in order to cover its monitoring costs. The first order condition from the expected return maximisation of the entrepreneurs, subject to the bank contract, gives rise to one of the three key equations in the financial frictions block together with the evolution of the net worth of entrepreneurs and the arbitrage equation for capital. The most important impact of the financial friction is that it 'accelatares' the impact of negative shocks, since the default risk increases during recessions, which has a negative impact on net worth and investment, that further rises the default risk as a consequence of the corporate bond spread.

The log-linearised equation, assuming a deterministic trend in productivity, that links the financial friction shock ε_t^{ω} and the expected spread is written as

$$\mathbb{E}_t\left[\tilde{R}_{t+1}^k - r_t\right] = \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c}\varepsilon_t^b + \varsigma_{sp,b}(q_t + \bar{k}_t - n_t) + \varepsilon_t^{\omega},$$

where ε_t^b is the risk premium shock, λ describes the habit formation on consumption, γ is the long-run growth rate, σ_c is the elasticity of intertemporal substitution. The transmission of the financial shock to aggregate investment via Tobin's q_t depends crucially of the parameter $\varsigma_{sp,b}$. If this parameter collapses to zero (in the absence of the financial friction shock ε_t^{ω}), the model is equivalent to one with no financial frictions. The financial friction shock follows an AR(1) process

$$\varepsilon_t^{\omega} = \rho_{\omega}\varepsilon_{t-1}^{\omega} + \sigma_{\omega}\eta_t^{\omega},$$

with variance σ_{ω}^2. This implies that the DSGE model with financial frictions has eight stochastic shocks. We are particularly interested in how the parameters $\varsigma_{sp,b}$, ρ_{ω} and σ_{ω} evolve over time since they have an impact on how the 'accelerator' mechanism, created by allowing for financial frictions, changes over time.

Our full set of measurement equations is described in Appendix 7.3. In addition to the seven observables employed by Smets and Wouters (2007), we add a time series of the corporate bond spread, Spread_t, measured as the difference between the BAA Corporate Bond Yield over the 10 Year Treasury Note Yield. This time series is linked to the financial friction block above by equation

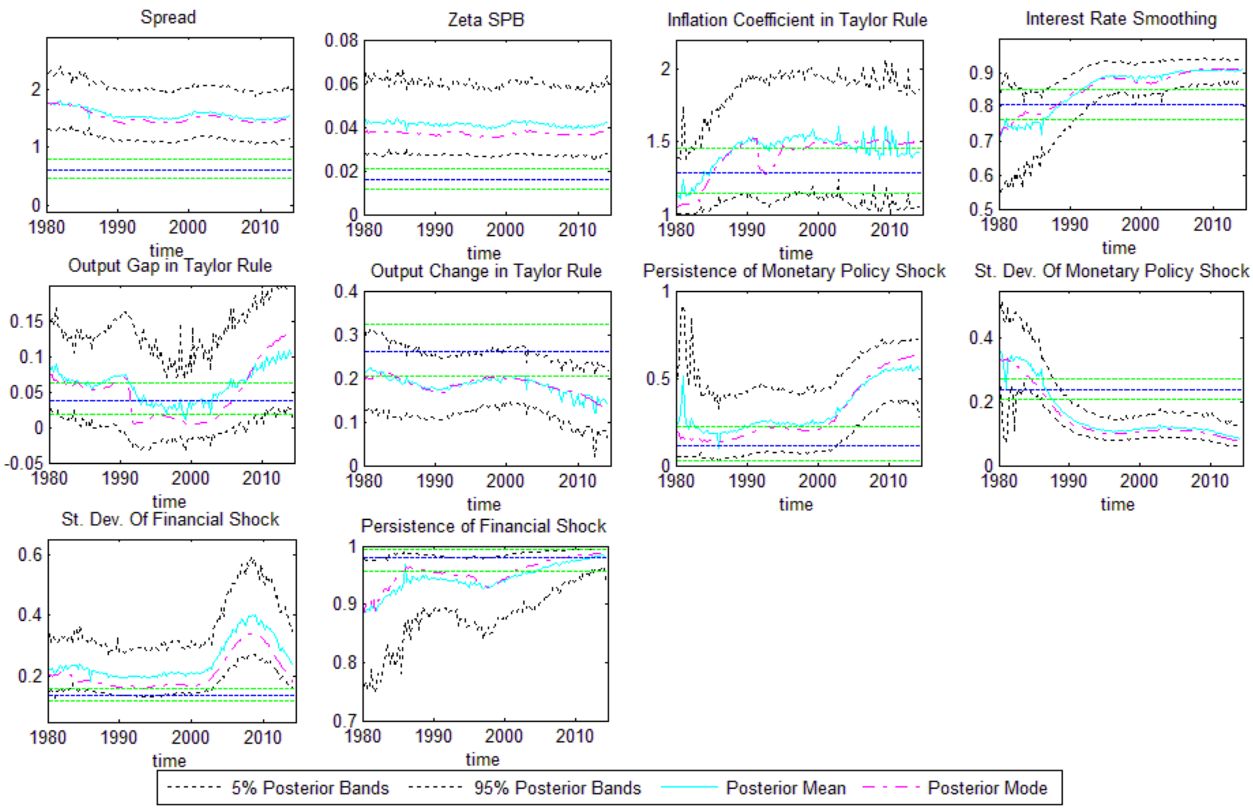
$$\text{Spread}_t = SP^* + 100 \times \mathbb{E}_t[\tilde{R}_{t+1}^k - r_t],$$

where r_t is the policy rate.

4.3 Empirical results

In this Section, we apply the quasi-Bayesian local likelihood (QBLL) method outlined in Chapter 3 to the DSGE model with financial frictions described in Section 4.2. We compare our results with the model estimated assuming fixed parameters. The QBLL method is applied with the weights w_{tj} generated by the normal kernel function and a bandwidth $\sqrt{T}$. The parameter prior distributions can be found in Appendix 7.3. These priors are the same as in Smets and Wouters (2007); for the financial friction block parameters we tried different prior specifications (see Appendix 7.3). The number of draws of the MH algorithm is 150,000, from which we drop the first 15,000. The scaling parameter for the MH has been adjusted in order to obtain rejection rates of 20%-30%[33]. We use U.S. data on eight observables described in Appendix 7.3 from 1970Q1 up to 2014Q2.

Figure 25: QBLL Estimates of DSGE model parameters with FF



Figures 26 and 27 present the estimates of selected parameters. The remaining parameters can be found in Appendix 7.3, and they are qualitatively similar to the estimates in Chapter 3. In Figures 26 and 27, the blue solid line is the posterior mean obtained by QBLL, with the black dotted lines displaying 5% and 95% posterior confidence intervals, and the pink dash-dotted line is the posterior mode obtained by QBLL. Finally, the dashed blue line is the posterior mean obtained by standard Bayesian methods with fixed coefficients, and the green dashed lines are the 5% and 95%

[33] Roberts et al. (1997) show that the optimal acceptance rate is 0.234 and their result serves as a rough benchmark in the literature; however, it is asymptotic and rests on the assumption that the elements of each chain are independent.

posterior quantiles. We would judge informally whether a parameter's variation is substantial by checking whether our estimates are outside the 5% and 95% quantiles of the posterior distribution of the fixed parameter model. We expect that the time-varying parameters, estimated by QBLL, will move slowly over time, in agreement with their variation representing stable and gradually changing relationships between the variables of the model, caused by smooth structural change. Parameters that vary in a erratic way would suggest that there exists no stable relationship between variables over time which might indicate model misspecification of a different nature to that arising out of smooth structural change.

Figure 26 displays the parameters of the Taylor rule, that is, interest rate smoothing and the relative impacts of inflation, output gap, and output growth on the policy rate. Estimated values are broadly in agreement with previous studies (Clarida et al. (2000), Cogley and Sargent (2002), Fernandez-Villaverde and Rubio-Ramirez (2008)) and suggest that the Federal Reserve has shifted the priority of its policy from output towards inflation in the mid-1980s. In particular, the Taylor rule inflation parameter starts increasing considerably especially after 1983, Paul Volcker's second term as a Chairman of the Federal Reserve, while the output gap coefficient falls during that period. Interest rate smoothing seems to have been low in the 1980s with tackling inflation being a priority and becomes higher through the second half of the sample. More interesting, we observe an increase in the output gap coefficient during the recent crisis, providing evidence that U.S. monetary authorities shifted attention to the sharply declining output. Monetary policy shock becomes more persistent during the crisis with interest rates near the Zero Lower Bound.

Figure 26 also includes the parameters of the financial friction block. Both the measurement equation parameter, SP^{*}, and the coefficient that measures the impact of financial frictions on Tobin's q,$\varsigma_{sp,b}$, have posterior means obtained with QBLL that are larger than assuming fixed parameters, but their values are in general stable over time[34]. The parameter measuring the persistence of the financial friction shock obtained with QBLL is in line with the one obtained with the fixed coefficient model since the QBLL estimate has large posterior confidence bands. In contrast, the volatility of financial frictions shocks increases twofold in period between 2007 and 2011 in comparison with the previous period and also with the fixed parameter posterior estimates.

[34] In the presence of time variation, fixed parameter estimation is inconsistent and hence results can be very different under both schemes especially in a non-linear setup. To address the question of why the confidence intervals of SP^{*} and $\zeta_{sp,b}$ obtained by BLL do not overlap with the fixed parameter ones, we offer a different explanation. Since the BLL approach uses a smaller sample (taking into account the down-weighting), it gives larger relative weight to the prior. For most parameters, this makes little difference, as the priors are not very tight. However, for SP^{*} and $\zeta_{sp,b}$, the BLL approach delivers estimates much more narrowly concentrated around the prior means than the fixed parameter model which uses a larger sample.

One might worry that this large increase may be caused by the inadequacy of the DSGE model to fit the data during the financial crisis period.

Figure 26: QBLL Estimates of DSGE model parameters with FF

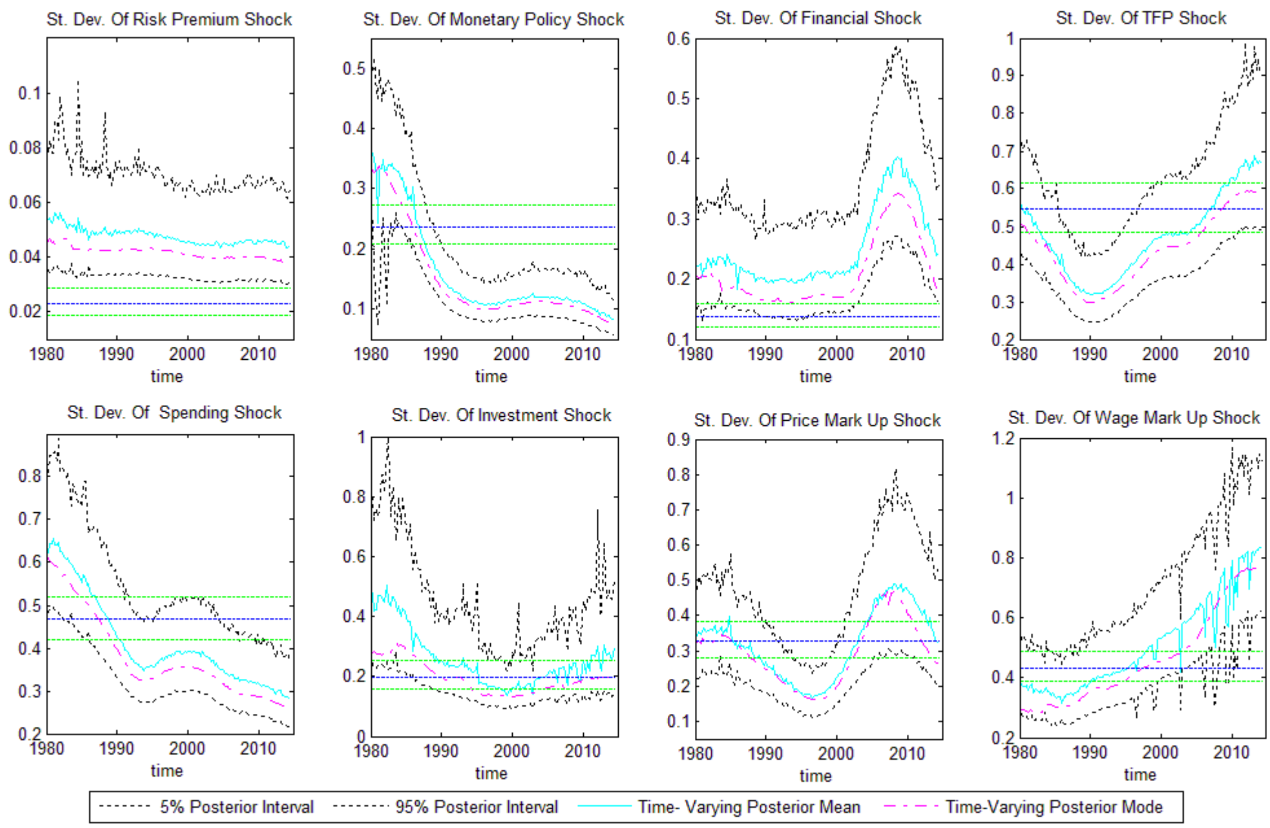


Figure 27 compares the time variation of the financial friction volatility with the volatilities of the other seven shocks. The results are broadly consistent with findings of low volatility during the Great Moderation (e.g. Primiceri (2005) and Sims and Zha (2006)). In particular, the volatilities of all shocks (except the wage mark up shock) fall in the late 1980s and remain low throughout the 1990s. The standard deviation of the monetary policy shock, for instance, is twice as large in the 1980s than in the 1990s. The volatilities of productivity and spending shocks are also small during the 1985-2005 period. These results suggest that the QBLL approach applied to the DSGE with financial frictions is able to reproduce previously documented changes in the variance of the shocks. Our new finding, however, is the clear increase in the financial friction shock variance from 0.2 to 0.4 during the 2007-2011 period. Similar relative size increases are not found for the other shocks since the total factor productivity shock only slightly exceeds the fixed parameter estimate during the most recent period.

In summary, the application of the QBLL approach to the DSGE model with financial frictions suggests that the volatility of the financial shock increased in the 2007-2011 but was small in the previous period. This adds a 'Good Luck' component to the interpretation of the Great Moderation period (1985-2006) since previous papers (Gali and Gambetti (2009), Benati and Surico (2009), Sims and Zha (2006)) looked at DSGE models that did not include a financial sector, and as consequence had no financial shocks. Moreover, we find that the volatility of the financial friction shock starts falling in 2012 and returns to pre-crisis levels in the end of 2014, suggesting a recovery of the economy from the financial crisis. An alternative explanation for the uncovered variation in the financial volatility is that the DSGE model we consider is too stylised and cannot capture fully the linkages between the financial sector and the rest of the economy and consequently, the impact of the events of the financial crisis appear in the variance of the financial friction shock in our empirical investigation.

4.3.1 Robustness Checks

In this Section, we report a number of robustness checks we performed in order to test the validity of the results presented in the previous Section. First, we checked the robustness of our findings by trying different prior specifications (see Appendix 7.3 for details) and by changing the trend assumption on productivity from deterministic to stochastic[35]. In all these specifications, we confirmed the results presented in Figure 1 and 2.

In addition, in Figure 28, we provide a comparison of the QBLL estimates of selected parameters[36] with ones generated with a simple rolling window scheme (solid green line)[37]. It is clear from Figure 28 that while the general pattern of the parameters does not change, the estimates obtained using the rolling window are considerably noisier. This is the case as at each point, a new observation is added and another one is thrown away. On the other hand, the QBLL, due to its capacity to reweight past observations without completely discarding any information, delivers smoother time variation. Noisy time variation in the DSGE parameters is not desirable for at least two reasons. First, if moving one observation forward causes large shifts in the values of some

[35] These additional results are available upon request.

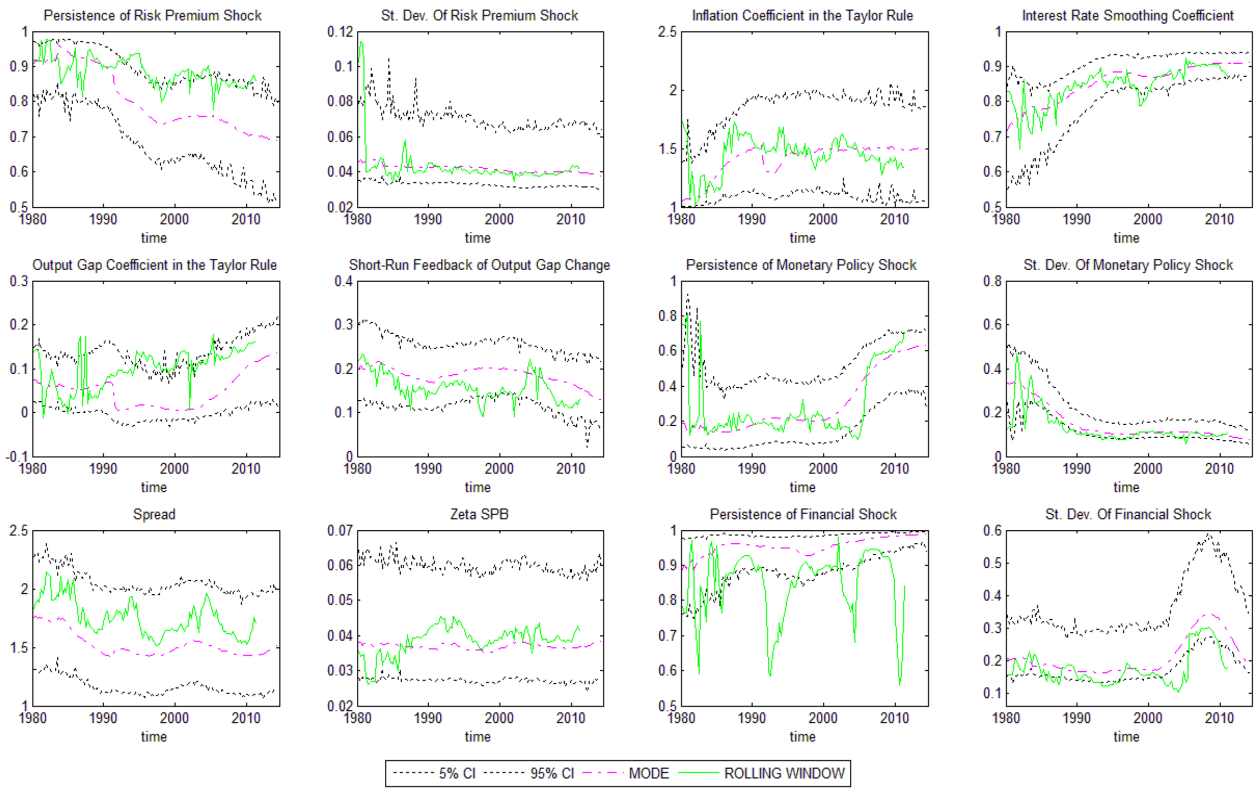
[36] The remaining parameters can be found in Appendix 6.5.

[37] For computational considerations, we only present the posterior mode estimates. To make the results comparable with the BLL results from the previous section, we use window size of $[2H+1]$ observations, where H is the bandwidth used for the normal kernel.

parameters, this might distort forecasting performance. Second, as argued in the previous Section, we believe that the variation in the DSGE parameters should be gradual, because it implies stable and gradually changing relationships between the variables of the model. The normal kernel has been found in the Monte Carlo study of Giraitis et al. (2014) to provide estimators with lower MSE compared to the flat kernel.

Furthermore, in order to assess the robustness of our results with respect to different spread variables, we estimated the model using the difference between the BAA corporate bond yield and the Fed Funds rate (solid blue line). Figure 29 displays the posterior modes of selected parameter estimates[38]. We discover that the results with this alternative spread specification do not alter our main conclusions.

Figure 27: Robustness Check



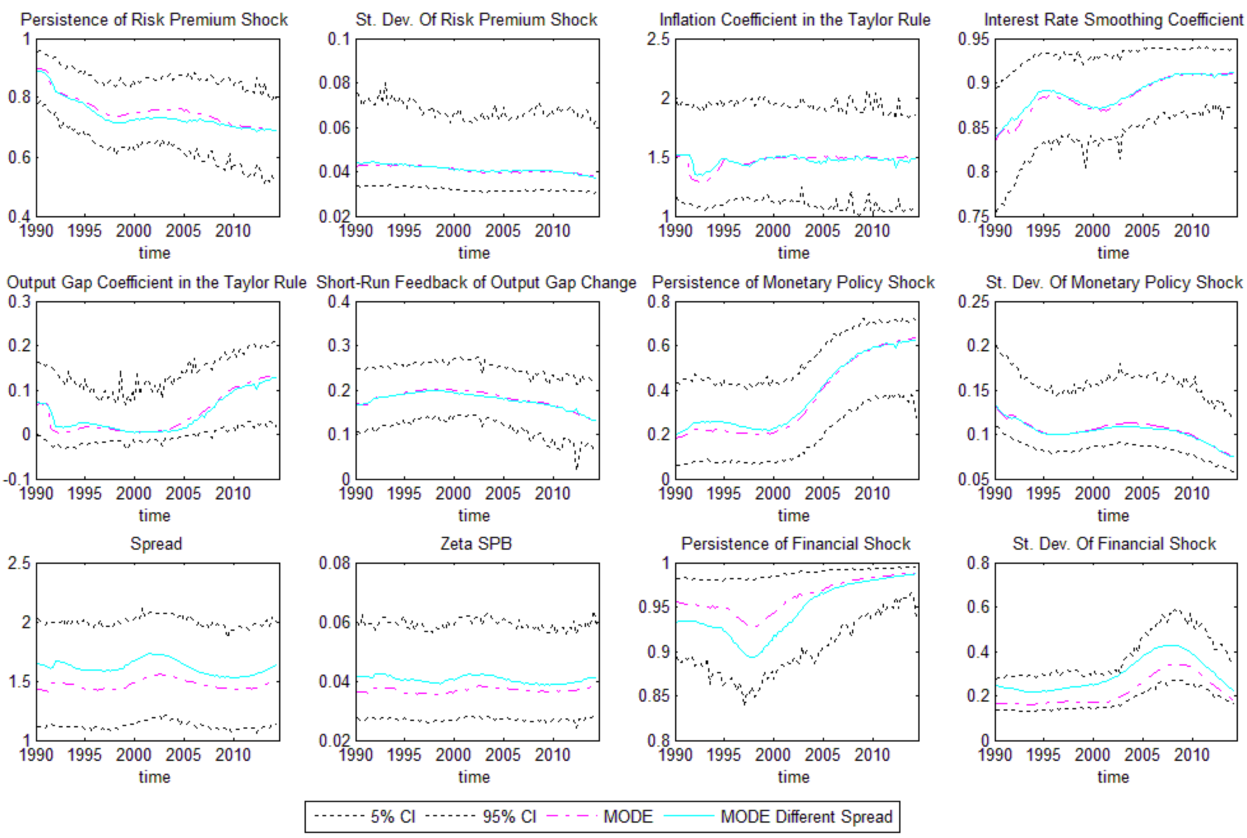
In addition, we also ran a small simulation exercise[39] in order to check if the QBLL approach works

[38] The remaining parameters can be found in Appendix 6.6.

[39] Due to computational time considerations, we only ran 10 replications, each with a sample size of 1000.

even in the absence of parameter time variation. We generated data from a fixed parameter DSGE model with financial frictions (dotted green line), using as a parameter vector the prior means[40]. Then, we applied our QBLL approach (solid blue line) to these artificial data. Figure 30 displays the resulting estimates from a representative replication for selected parameters[41] and demonstrates how the QBLL approach recovers the true parameters with virtually no time variation. This suggests that the method is valid even in the absense of time variation and therefore the uncovered variation in the model's parameters in our empirical application is not spurious but is instead a feature of the US data used for estimation.

Figure 28: Robustness Check

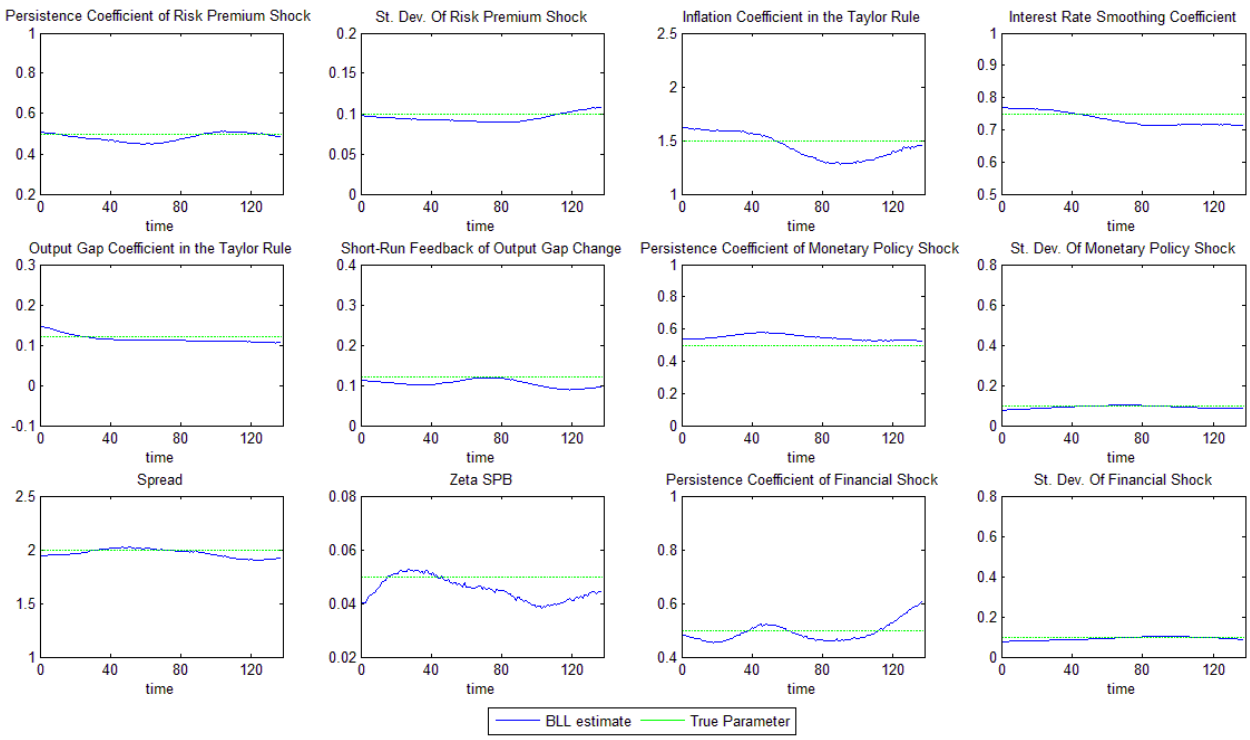


Finally, our choice of bandwidth in the empirical application in the previous Section is motivated by the optimal bandwidth choice used for inference in time varying random coefficient models

[40] We set the standard deviations to 0.1, as the prior mean for these is infinite.
[41] The remaining parameters can be found in Appendix 6.7.

in Giraitis et al. (2014). In addition, in Chapter 2 and 3, we performed a number of robustness checks with respect to different values of H. Moreover, in an application to the Smets and Wouters (2007) model without financial frictions, Chapter 3 showed that $H = \sqrt{T}$ delivers the best forecast performance for most variables.

Figure 29: Simulation Exercise

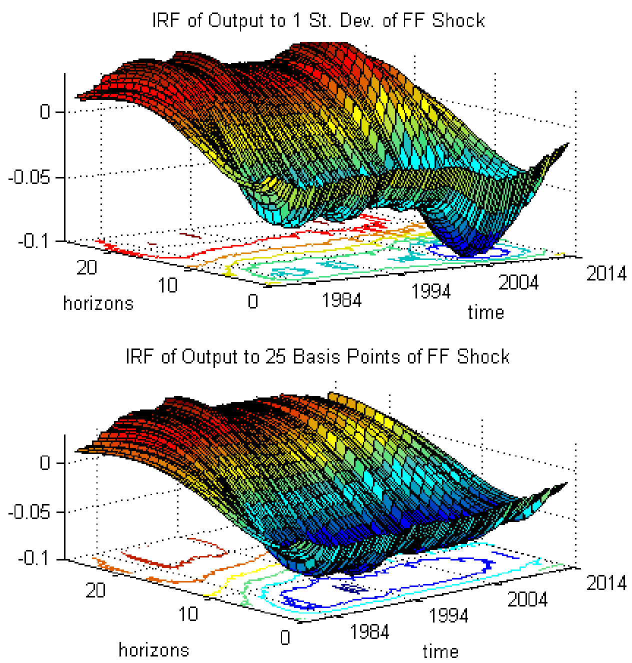


4.3.2 Time-varying impulse response functions

The main objective of this subsection is to evaluate how the financial frictions shock propagates to the rest of the economy over time. Our previous results suggest that the size of the financial shock is larger in the 2007-2011 period. However, the parameter that governs the transmission of the shock to macro variables, $\zeta_{sp,b}$, does not change very much. Figure 31 displays the impulse response functions of output and investment. The top panel of Figure 31 describes the responses to one-standard deviation of the shock, so it captures the effect of the shock on the desired variables over time while also taking into account its changing size. The bottom panel of Figure 31 describes the responses to 25-basis-points shock, which are useful for investigating the changes in the transmission while keeping the size of the shock constant over time.

We can see that the negative responses of output 10 quarters after the shock are 0.1% during 2008-2011 period instead of 0.05% prior to 2008. Similarly, investment, which is the main channel through which the financial shock affects output, responds much more sharply during the 2007-2011 period, with an accumulated response of minus 6.5% in the five years after the shock hits, as supposed to minus 1.5% in the pre-crisis period. If we consider the impact of a fixed-sized shock instead, the resulting responses of both output and investment are virtually the same across periods. This confirms our conjecture that what has changed over time is the size of the financial shocks rather than the way in which financial markets operate in the model.

Figure 30: Responses to 1 st. dev. and 25 basis points of financial friction shock



4.3.3 Forecasting

Our previous results indicate that the QBLL approach applied to the DSGE model with financial frictions is able to capture important variation of the parameters over time. In particular, Figure 27 provides exhaustive evidence of changes in the volatility of the shocks. The literature on forecasting with time-varying volatilities (e.g., Carriero, Clark and Marcellino (2015b)) suggests that we should expect improvements in forecasting accuracy in particularly when evaluating the predictive densities. In this subsection, we use the algorithm outlined in Chapter 2 to generate density forecasts for the observables, using the posterior distribution of the parameters at the last period T of the in-sample to generate the out-of-sample predictions. Our forecast origins are 2000Q1-2012Q2 and we generate projections 1 to 8 quarters ahead. In addition to our time-varying DSGE model with financial frictions (TV FF), we compute forecasts for the DSGE model with (fixed FF) and without financial frictions assuming fixed parameters. The standard Smets and Wouters (2007) (SW) model has been evaluated by Edge and Guerkaynak (2010), Woulters (2012) and Del Negro and Schorfheide (2013b), and it is able to perform well at long forecast horizons for output growth and inflation, so we use it as a benchmark in Table 17.

	2000Q1-2006Q4				2007Q1-2012Q2			
	TV FF relative to SW				TV FF relative to SW			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	1.00	1.20	1.26	1.27	0.90**	0.90*	0.95	0.97
Investment Growth	0.87	0.74*	0.68*	0.88	0.95	0.99	1.12	1.17
Inflation	0.99	0.98	0.97	0.90	1.03	1.05	1.05	0.93
Fed Funds Rate	0.89**	0.89**	0.86**	0.79	0.64**	0.66*	0.86	1.05
	Fixed FF relative to SW				Fixed FF relative to SW			
Output Growth	1.21	1.53	1.77	1.65	0.92	0.86*	0.86**	0.84*
Investment Growth	0.93	0.84	0.80	0.95	0.94	0.93	1.06	1.19
Inflation	1.02	1.05	1.09	1.01	0.98*	0.98*	0.99	1.10
Fed Funds Rate	0.82**	0.79**	0.84*	0.89	0.97	0.95	1.00	0.98

Table 17: RMSFEs. The table reports ratios of RMSFEs relative to the SW model RMSFEs. '*', '**' and '***' indicate rejection of the null of equal performance against the one-sided alternative at 10%, 5% and 1% respectively, using a Diebold and Mariano test.

Table 17 evaluates the performance of point forecasts using root mean squared forecast errors (RMSFEs) for output growth, inflation and the Fed Funds rate, since these variables are of prime interest. In addition, we also report the forecasts for investment growth[42] as it is the channel through which the financial markets enter the model[43]. Entries are ratios with respect to the SW model benchmark. Values smaller than one imply that the model (either the TV FF or fixed FF) is more accurate than the benchmark. Table 18 presents the relative forecasting performance in

[42]The forecasts for the remaining variables can be found in Appendix 6.4. They lead to qualitatively similar conclusions.

[43]A forecasting comparison with an AR(1) and a TVP AR(1) models can be found in Appendix 6.4. The autoregressive models are included because it is important to verify that the DSGE model is at least as accurate as univariate statistical models.

terms of log predictive scores. The log score is computed as the value of the predictive density evaluated at the realised target variable and is therefore a measure of the precision of the density forecasts. We test whether a model is statistically more accurate than the SW benchmark with the Diebold and Mariano (1995) statistic computed with Newey West estimator for the standard errors. One, two and three stars indicate rejection of the null of equal performance against the one-sided alternative of better performance over the SW benchmark at 10%, 5% and 1% respectively. Both tables present results for two sub-periods: 2000Q1-2006Q4 and 2007Q1-2012Q2. The first period is relatively tranquil in comparison with the second one.

	2000Q1-2006Q4				2007Q1-2012Q2			
	Density TV FF relative to SW				Density TV FF relative to SW			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	0.13***	0.10*	0.10*	0.05	0.05	0.05	-0.09	-0.09
Investment Growth	0.19*	0.29*	0.35*	0.11	-0.04	0.31	-0.27	-0.22
Inflation	-0.08	-0.07	0.03	0.11	5.56	3.60	2.52	-0.36
Fed Funds Rate	0.49***	0.18	-0.16	-0.12	0.47***	0.31	-0.08	-0.48
	Density Fixed FF relative to SW				Density Fixed FF relative to SW			
Output Growth	-0.10	-0.15	-0.19	-0.19	0.01	0.08	0.12*	0.05
Investment Growth	0.01	0.01	0.01	-0.06	0.16	0.52	0.14	-0.11
Inflation	0.00	-0.04	-0.11	-0.09	-2.92	0.59	1.58	-0.09
Fed Funds Rate	0.03**	-0.02	-0.08	-0.05	0.02	0.01	-0.03	0.11

Table 18: Log Scores. The table reports differences of log predictive scores from to the SW model log scores. ‘*’, ‘**’ and ‘***’ indicate rejection of the null of equal performance against the one-sided alternative at 10%, 5% and 1% respectively, using a Diebold and Mariano test.

Del Negro, Hasegawa and Schorfheide (2014) and Kolasa and Rubaszek (2015) documented that the DSGE model with financial frictions does not improve forecasts in comparison with the Smets and Wouters (2007) model in the period before 2007. The results in Table 17 confirm this since the inclusion of financial frictions worsens the forecasts of inflation, output and investment growth in the period 2000-2006, while improving forecasts of the interest rate. The TV FF model brings the forecasting performance at similar levels to the SW model during this tranquil period. In addition, during the volatile period of 2007-2012, the TV FF model confirms previous findings of relatively good performance of financial friction models in comparison to the standard SW model. For the Fed Funds Rate, the TV FF model delivers statistically significant improvements over the SW model for both point and density forecasts. One explanation is that by allowing for time variation in the coefficients of the Taylor rule, we obtain a value for the smoothing parameter that is close to one; that is, the forecasts from the Taylor rule resemble random walk forecasts, which is an adequate model for forecasting the Fed Funds Rate in the vicinity of the Zero Lower Bound. Table 18 presents similar results for density forecast performance for selected variables[44] using log scores. Qualitatively, the results are similar to the ones using RMSFEs.

4.4 Summary

This Chapter employs the quasi-Bayesian Local Likelihood approach developed previously in Chapter 2 and 3 to a DSGE model that combines the Smets and Wouters (2007) model with financial frictions as in Bernanke et al. (1999). As a consequence, this Chapter proposes a time varying DSGE model with financial frictions. Our results suggest that the parameter governing how financial friction shocks affect investment decisions is stable over time, but the volatility of the financial shock jumps in the period 2007-2012 and returns to the pre-crisis values after 2012. Moreover, when looking at the impulse response functions, we find that the responses of output and investment to 25 basis points of the financial shock do not change over time. In contrast, when we consider the responses to a standard deviation of the shock, taking into account the changing volatility over time, we observe a substantial change in the way these variables respond to financial shocks during the period of the recent crisis. This evidence leads us to provide an interpretation of the recent financial crisis as a 'Bad Luck' event , that is, it is caused by changes in the volatility of financial shocks while taking into account policy changes. An alternative explanation of the results presented in this Chapter is that the DSGE model we consider is perhaps too stylised to fully account for the connection between the financial sector and the macroeconomy and, as a consequence, the events of the 2008 crisis appear in the variance of the financial friction shock instead.

Finally, our forecasting exercise demonstrates that the time varying model with financial frictions improves the forecasting performance of the financial friction model especially in the tranquil 2000-2006 period.

[44] The density forecast for the remaining variables, as well as density forecast comparison with an AR(1) and a TVP AR(1) models can be found in Appendix 6.4.

5 Time Varying COMPASS Model of the U.K. Economy

5.1 Introduction

As argued in Chapter 3 and 4, medium-sized DSGE models, such as the one presented in Smets and Wouters (2007), are routinely estimated by academics and policy makers for the U.S. economy. However, such models are tailor-made for the U.S. and the standard assumption made is closed economy, which might not be appropriate for small open economies. Hence, central banks in other countries have developed their own DSGE models aimed at fitting better data in their economy, e.g. the Riksbank's RAMSES model (Adolfson et al. (2007)) or the ECB's New Area Wide model (Christoffel et al. (2010)). A medium-sized DSGE for the U.K. economy, referred to as COMPASS, was developed by Burgess, Fernandez-Corugedo, Groth, Harrison, Monti, Theodoridis and Waldron (2013). COMPASS is an open economy New Keynesian model which shares many theoretical features of other DSGE models.

As discussed in Chapter 3, the estimation of DSGE models relies on the assumption that the model's parameters are not subject to change; i.e. parameters are structural in the Lucas sense - invariant to both policy and shocks. In general, this is a strong assumption that is worthy of testing. And, over particular time periods, there are reasons to doubt it will hold. For example, the UK economy is likely to have experienced several structural breaks over several decades associated with changes in monetary regime. Following the end of the Bretton Woods system and various crises in the mid 1970s, UK monetary policy was set by the government with the aim of controlling various different monetary aggregates. This framework was replaced in the late 1980s by informal exchange rate targeting, which was formalised in 1989 when the UK entered the Exchange Rate Mechanism. The ERM came to an end in 1992 when the target level of sterling became unsustainable, after which it was replaced by inflation targeting, which was conducted by the Government until 1997 and the Monetary Policy Committee of the Bank of England subsequently. Putting aside the certainty that parameters describing the behaviour of monetary policy would have changed over this period, it is hard to believe that such fundamental changes in monetary regime would not have been accompanied by changes in, for example, parameters describing the price setting behaviour of firms. More recently, the Great Recession of 2008 caused a large and persistent decline in activity relative to pre-crisis levels, which is hard to reconcile with pre-crisis parameter estimates. And, once the Bank rate had become constrained by an effective lower bound, the picture is further complicated by the adoption of large scale asset purchases and an increased use of forward guidance.

At the very least, it is hard to believe that parameters governing macroeconomic volatility (the standard deviations of shocks) would have been unaffected by the crisis. Such parameter variation, if ignored can result into misspecification and poor model fit. Hence, even if one believes in the structural nature of DSGE parameters, it is important that one recognises at least the possibility of parameter drift. The estimation exercise in Burgess et al. (2013) sidesteps some of these issues by restricting the estimation to the post inflation targeting, pre financial crisis sample (1993-2007) and by removing some of the remaining trends from the data (like the disinflation following the adoption of inflation targeting in the early 1990s). This strategy has the advantage of making it more likely that the constant parameter assumption is valid, but has two main drawbacks. The sample is, by almost any standard, relatively short. Although this is mitigated by the use of Bayesian techniques, it is certainly not ideal, because there would be additional variation over a longer sample that should improve the identification of the posterior parameter estimates. More importantly, in hindsight, it is clear that the 1993-2007 sample represents an extraordinarily benign period of steady growth and broader macroeconomic stability. Parameter estimates over this period are, therefore, unlikely to be a good guide to the future, which has obvious implications for forecasting. Time variation in the model's parameters can also be a signal for misspecification and hence allowing for time variation can improve the fit of the model as well as serve as a future guide to amend and revise it.

In this Chapter, we employ the approach developed in Chapter 3 and apply it to COMPASS to investigate the structural nature of the parameters of the model. The flexibility of the approach in the face of structural change permits the estimation of COMPASS over a longer period, alleviating the need to restrict the sample to post-1993 and pre-crisis data. In this version of this Chapter, we investigate the time-variation that is due to the financial crisis as well as the period 1987-1993, leaving estimation using pre-1987 data for future work[45]. Our prioritisation of possible time variation induced by the financial crisis reflects that it is more immediately applicable to. In addition, we investigate the time-variation that is due to the financial crisis, in part because that is more immediately applicable to analysis of the current economic outlook. The evident changes in the transmission of monetary-policy shocks and in the estimated productivity trend we find, as well as the large changes in the volatility parameters, highlight the importance of this exercise and contribute to the small but expanding literature on estimating DSGE models with time variation in the parameters. Our main empirical finding is that the risk premium shock in the model has

[45]In a more recent version of this work, we go back to 1975 and handle the missing observations for some series early in the sample with the Kalman filter.

played an important role in the UK economy during the recent financial crisis. In particular, we find the volatility of the shock is nearly twofold its 2000 level and the transmission of the shock to economic variables, such as the nominal interest rate, has changed considerably during the crisis. Finally, since COMPASS is used routinely for generating quarterly forecasts for the UK economy, allowing the model's parameters to change and using only their most recent values for generating predictions is a useful modification that makes the model more flexible and delivers some gains in the model's forecasting performance[46].

The remainder of the Chapter is organised as follows. Section 5.2 presents COMPASS, Sections 5.4, 5.5, 5.6 and 5.7 contains the empirical results and forecasting comparison and Section 5.8 concludes.

5.2 Model and Data

We apply the methodology developed in Chapter 3 to an operational medium size DSGE model of the UK economy, namely that used as the main organizing framework in the Bank of England forecasting process, since late 2011. A detailed description of the model - known as COMPASS, which is short for Central Organizing Model for Projection Analysis and Scenario Simulation - is presented in Burgess et al. (2013).

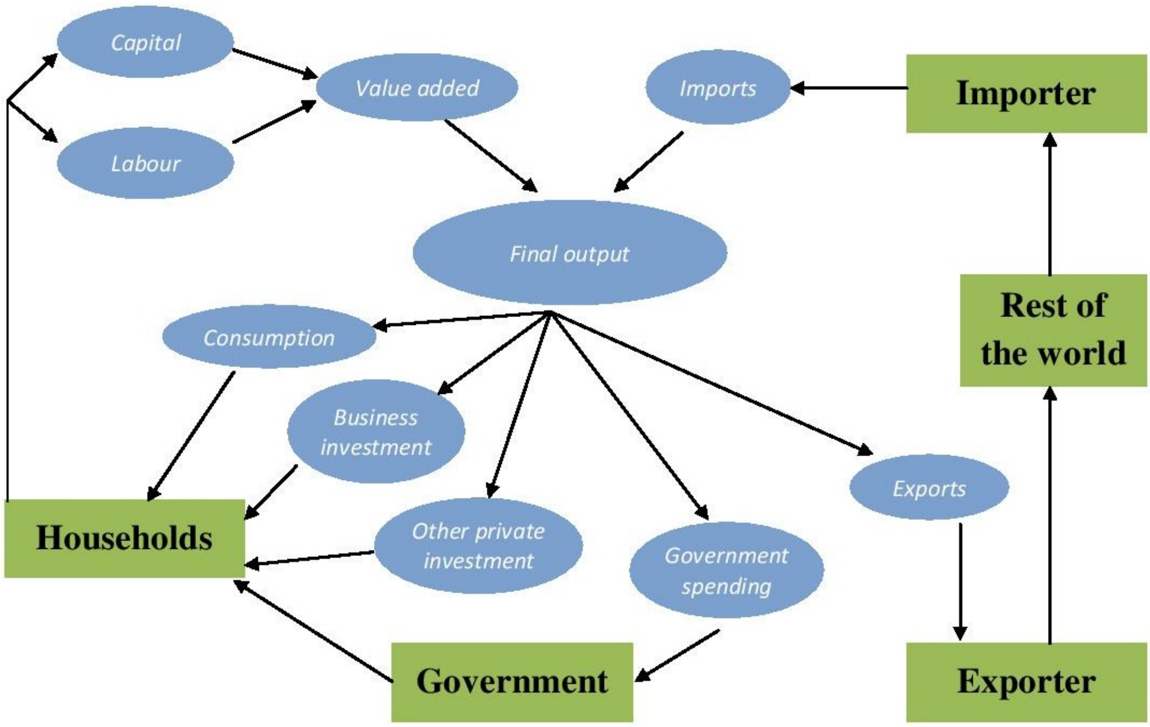
For our purposes, a high level description of the model setup and an overview of the key mechanisms is enough. In particular, for the most part we will report log-linearized conditions, referring to Burgess et al. (2013) for the derivation from first principles. The economy is made up of five main economic agent types: households, firms, the monetary policy maker, the government and the rest of the world. We will briefly describe each of them in turn.

5.2.1 Households

Households come in two types: optimizing and hand-to-mouth. Optimizing households make the following key economic decisions:

[46] The advantage of the time varying version of the model over the standard fixed parameter version is its better forecast performance, which comes at no extra computational cost since for generating forecasts, only the posterior distribution at the last period is required. However, one might still want to use the standard fixed parameter version in a policy setting: i) from a policy perspective it is not clear how useful the time varying version is because the parameter changes are external to agents in the model; ii) it is more computationally demanding to estimate, as it requires characterisation of the posterior density at each point in time.

Figure 31: Flow of Goods and Services in the COMPASS



1. *Intertemporal Consumption Decision*

$$c_t = \frac{1}{1+\psi_C+\epsilon_\beta(1-\psi_C)\epsilon_C^{-1}}\left[\mathbb{E}_t c_{t+1}+\psi_C c_{t-1}\right] - \frac{\omega_o(1-\psi_C)}{(1+\psi_C)\,\epsilon_C+\epsilon_\beta(1-\psi_C)}$$
$$\times\left[r_t-\mathbb{E}_t\pi^Z_{t+1}+\hat{\varepsilon}^B_t-\mathbb{E}_t\gamma^Z_{t+1}\right]+(1-\omega_o)\,\frac{wL}{C}\left[w_t+l_t-\frac{\mathbb{E}_t w_{t+1}+\mathbb{E}_t l_{t+1}+\psi_C\,(w_{t-1}+l_{t-1})}{1+\psi_C+\epsilon_\beta(1-\psi_C)\epsilon_C^{-1}}\right] \tag{39}$$

Equation (39) is a relatively standard consumption Euler equation. The first term on the RHS contains past consumption because agents' utility is subject to habit formation while the second term contains a risk-premium term $\hat{\varepsilon}^B_t$, which measures a wedge between the rate set by policy-makers and that faced by consumers. The third term on the RHS depends on the fact that a share $(1-\omega_0)$ of the agents are hand-to-mouth so their consumption depends on their labor income.

2. *Investment Decision*

Optimizing Households can smooth consumption over time by either investing in physical capital or investing in financial assets. Investment is subject to adjustment costs, which result in the following standard-looking Euler equation:

$$i_t = \frac{\beta\Gamma^H}{1+\beta\Gamma^H}\left(i_{t+1} + \gamma^Z_{t+1}\right) + \frac{1}{1+\beta\Gamma^H}\left(i_{t-1} - \gamma^Z_t\right) + \frac{1}{\left(1+\beta\Gamma^H\right)\left(\Gamma^H\Gamma^Z\Gamma^I\right)^2}\left(\frac{tq_t}{\psi_I} + \hat{\varepsilon}^I_t\right) \tag{40}$$

where tq_t is Tobin's q value of one unit of capital, which depends on the difference between the future expected streams of returns on capital r^K_t and the real interest rate, adjusted for the risk-premium shock:

$$tq_t = \frac{1-\delta_K}{r^K + (1-\delta_K)}\mathbb{E}_t tq_{t+1} - \left(r_t - \mathbb{E}_t\pi^Z_{t+1} + \hat{\varepsilon}^B_t\right) + \frac{r^K}{r^K + (1-\delta_K)}\mathbb{E}_t r^K_{t+1}$$

3. *Portfolio Decision*

Strictly speaking, households delegate their portfolio decision to risk-neutral portfolio packagers who collect deposits from households and buy domestic and foreign bonds. The end-result is the following UIP condition:

$$q_t = \mathbb{E}_t q_{t+1} + \left(r_t - \mathbb{E}_t\pi^Z_{t+1}\right) - \hat{\varepsilon}^{B^F}_t$$

which is an arbitrage condition between returns on domestic and foreign assets.

4. *Wage Setting*

Households supply differentiated labor services in a monopolistically competitive setting. As a result, they have some degree of wage-setting power, i.e. they set their nominal wage at a markup over their marginal-rate of substitution between consumption and leisure (see Erceg, Henderson and Levin (2000)). The wage setting process is also subject to an adjustment cost (Rotemberg (1982)) which, when allowing for indexation to the previous periods' wage rate governed by ξ^W, results in the following wage Phillips Curve:

$$\pi^W_t = \hat{\mu}^W_t + \frac{\hat{\varepsilon}^L_t + \epsilon_L l_t + \frac{\epsilon_C\left(c^o_t - \psi_C c^o_{t-1}\right)}{1-\psi_C} - w_t}{\phi_W\left(1+\beta\Gamma^H\xi^W\right)} + \frac{\xi^W}{1+\beta\Gamma^H\xi^W}\pi^W_{t-1} + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi^W}\mathbb{E}_t\pi^W_{t+1}$$

where the first term on the RHS is the markup, which is allowed to vary over time, and the second represents the marginal rate of substitution.

Hand-to-mouth households do not have access to financial markets by assumption. They will consume their labor income in every period and will receive government transfers to ensure their income grows in line with that of optimizing households along the balanced growth path.

5.2.2 Firms

The production sector in COMPASS is somewhat more complicated than in most medium-size DSGEs (e.g. Smets and Wouters, 2007) because of interactions with the rest of the world and because the model is required to provide a detailed breakdown of various "demand components".

1. *Value Added Producers*

   This is the most standard of sectors. Firms hire capital and labor, which are used in a Cobb-Douglas production function:

   $$v_t = (1-\alpha_L)\, k_{t-1} + \alpha_L l_t + \hat{\varepsilon}_t^{TFP}$$

   Firms face monopolistic competition, hence they set their price at a markup over their marginal cost. They are also subject to price-adjustment costs, which ultimately result in the following value-added inflation Phillips Curve:

   $$\pi_t^V = \hat{\mu}_t^V + \frac{1}{\phi_V\left(1+\beta\Gamma^H\xi_V\right)} mc_t^V + \frac{\xi_V}{1+\beta\Gamma^H\xi_V}\pi_{t-1}^V + \frac{\beta\Gamma^H}{1+\beta\Gamma^H\xi_V}\mathbb{E}_t\pi_{t+1}^V \qquad (41)$$

2. *Importers*

   They simply buy goods and services from the rest of the world and sell it domestically. They set prices in domestic currency at a markup over the marginal cost.

3. *Final Output Producers*

   They operate a similar technology and are subject to the same time of pricing frictions as the value added producers. However, their production function combines value-added output and imports to produce final output.

4. *Retailers*

   Retailers operate in a competitive market and transform final output into Consumption, Business and Other Investment, Government Spending and Exports, as Figure 31 illustrates. In doing so, they operate linear technologies which differ in their productivities. This is a technical expedient to accommodate different trend growth rates in the corresponding observable variables.

5. *Exporters*

   They buy export goods from the corresponding retail sector and sell it to the rest of the world. They operate in a monopolistically-competitive market subject to price-adjustment frictions, which results in a Phillips Curve along the lines of equation (41) (prices are set in foreign currency).

5.2.3 Monetary Policy

In COMPASS policy rates are set according to a simple linear reaction function:

$$r_t = \theta_R r_{t-1} + (1 - \theta_R) \left[\theta_\Pi \left(\frac{1}{4} \sum_{j=0}^{3} \pi^Z_{t-j} \right) + \theta_Y \hat{y}_t \right] + \hat{\varepsilon}^R_t$$

which feature a response to annual inflation in deviation from its target (which also corresponds to its steady-state level in COMPASS), the output gap and a degree of interest-rate smoothing governed by θ_R. The Taylor rule is not subject to inflation-target shocks (see Del Negro et al. (2014)) because the target has never changed in the UK since it was first introduced[47]

5.2.4 Government Spending

Real-government spending, in deviations from trend, is assumed to follow a simple autoregressive process:

$$g_t - g_{t-1} + \gamma^Z_t = (\rho_G - 1)\, g_{t-1} + \hat{\varepsilon}^G_t \tag{42}$$

where γ^Z_t measures labor-augmenting productivity and spending is financed via lump-sum taxes on optimizing households.

5.2.5 Rest of the World

While it still retains some features of a closed economy, e.g. the interest-rate setting power, the UK economy is, in many ways, modeled as a small open economy.

[47] The data is detrended in a way that we will details in the next section so that this is not an issue even prior to the introduction of the 2 percent inflation target.

In particular, this implies that world output and prices are independent of domestic shocks, with one important exception, which is necessary for balanced growth, namely the fact that the world economy inherits the domestic permanent labor productivity shock according to a term ω_t^F which ensures the catching up of the world to the domestic productivity shock does not happen instantaneously.

As a result, the world economy is described by three simple equations:

$$z_t^F = \omega_t^F + \rho_{Z^F} z_{t-1}^F + \hat{\varepsilon}_t^{Z^F} \tag{43}$$

$$p_t^{X^F} = \rho_{PX^F} p_{t-1}^{X^F} + \hat{\varepsilon}_t^{PX^F} \tag{44}$$

$$x_t = z_t^F + \hat{\varepsilon}_t^{\kappa^F} - \epsilon_F \left(p_t^{EXP} - p_t^{X^F} \right) \tag{45}$$

which describe world output (which includes the ω_t^F term described above), world prices and the equation governing the demand for UK exports, which is an increasing function of world output and a decreasing function of the prices of UK exports (p_t^{EXP}) relative to world prices, the ε terms representing exogenous disturbances.

5.3 Data

The model is estimated using fifteen macroeconomic quarterly time series[48] for the period of 1987Q3 to 2012Q3. The variables, data transformations and measurement equations are described in Appendix 7.4. All variables, except for the policy rate, are log-differenced. We also subtract what we call time-varying trends from some of the variables, e.g. exports. This terms represent a simple way to capture the fact while the model implies a common long-run growth rate for all real variables, the average growth rates measured in the sample differ. Most time-varying trends capture this idea, i.e. they subtract a constant from the log-differenced variable so that it roughly averages the same growth rate as UK GDP.

The exception to this rule in the inflation time-varying trend, whose purpose is conceptually slightly different. Its purpose is to correct for the fact that prior to 1992 there was no explicit inflation target so there should be no expectation that inflation hovered around 2 percent on average. Hence we pre-process the data so that imposing a 2 percent average for inflation throughout is not restrictive. Irrespective of the motivations behind these corrections, what matters for our purposes

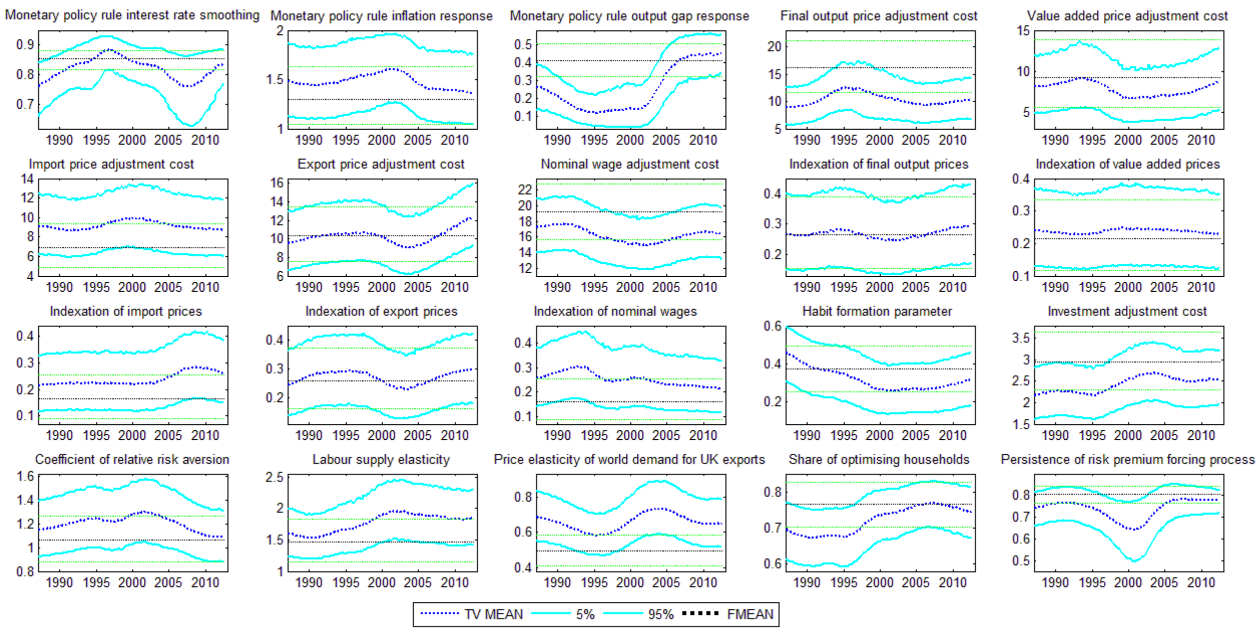
[48]Notice that COMPASS features 16 structural shocks and 7 measurement errors, so the number of shocks is larger than the number of observables.

is that they are applied before the data is shown to the model and they do not depend on model estimates. Rather they are part of the transformation of raw data into model-consistent observables and capture features of the data COMPASS is not meant to model accurately (e.g. secular trends in world trade).

5.4 Estimation Results

In this Section we present results for the fixed parameter model and also for the version with time varying parameters estimated with the QBLL method described in Chapter 3. In both cases, we employ most priors from Burgess et al. (2013)[49], with four MH chains with 220,000 draws each, from which we drop the first 20,000, and apply a thinning rate of 50%[50]. We set the MH scaling parameters such that acceptance rates are around 25%. We apply the QBLL method using the Normal kernel function in equation (5) with a bandwidth size of $\sqrt{T}$. The results are presented

Figure 32: QBLL Estimates COMPASS

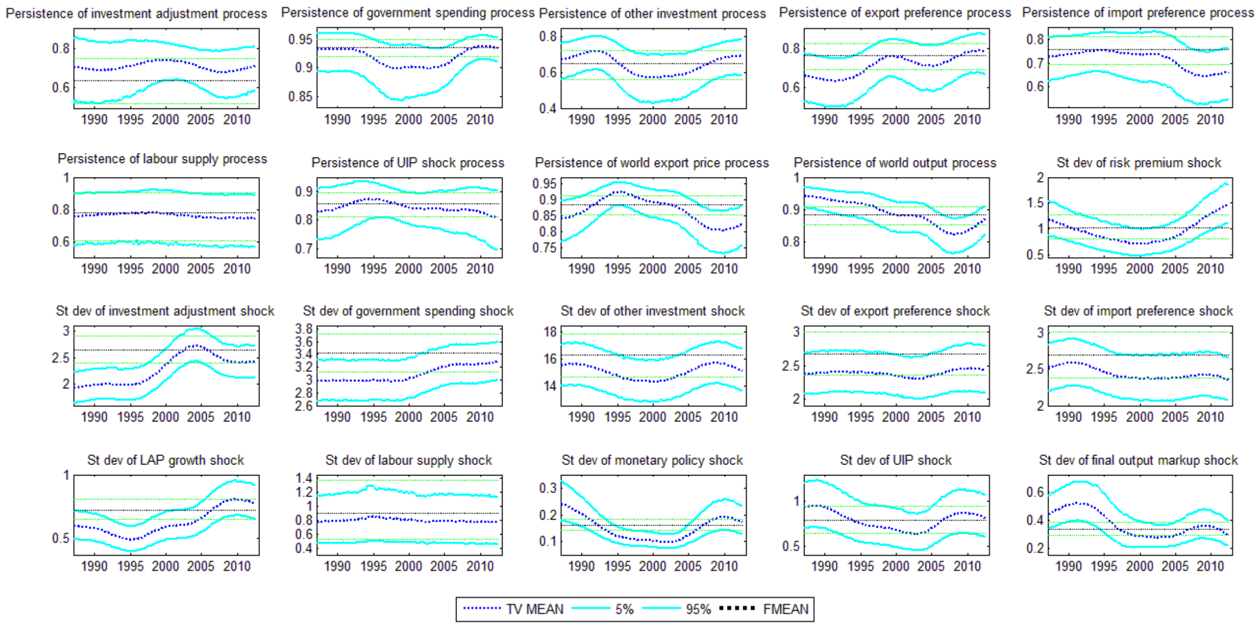


in Figures 33-35, where the dark blue dotted line is the posterior mean obtained with the QBLL approach, the light blue dotted lines are the 5% and 95% posterior bands, the dotted black line is the posterior mean obtained with standard fixed parameter Bayesian estimation and the dotted

[49] See Appendix for details.

[50] That is, our effective number of draws is 400,000 after thinning and burning.

Figure 33: QBLL Estimates COMPASS



green lines are the corresponding 5% and 95% posterior bands around it. We employ Figures 33-35 to judge informally whether a parameter's variation is substantial by checking whether the QBLL estimates are outside the confidence bands of the fixed-parameter estimate.

The Taylor rule parameters are subjected to some time variation. In particular, the output gap coefficient increases twofold from its early 2000s level, indicative of a possible shift in the weight attached to stabilising output versus headline inflation in response to oil price shocks and the financial crisis. At the same time, the interest rate smoothing parameter – the coefficient on the lagged policy rate in the Taylor rule – declines, indicative of a desire for the central bank to react more quickly to unfolding events than the reduced-form Taylor rule characterisation of their past behaviour would have prescribed. In addition, consistent with the substantial easing of policy during the crisis, the volatility of the monetary policy shock increases towards the end of the sample. In fact, we find considerable time variation in some parameters that govern the stochastic processes of the exogenous shocks. In particular, we find that both the persistence and the volatility of the risk premium shock display a U-shape over our sample period, with its standard deviation more than doubling in the end of the sample compared to pre-crisis levels. Moreover the volatility of the world output shock also increase in the end of the sample, in order to capture the increased volatility we observe in foreign output data during the crisis. Trend productivity,

Figure 34: QBLL Estimates COMPASS

the steady state growth rate of final output per capita, also falls slightly during the crisis. The standard fixed parameter approach underestimates this value between 1995-2005 but overestimates it at the beginning and end of the sample. This parameter is important for generating forecasts[51] as it appears in several of the measurement equations (output, consumptions, investment, export, imports and wages) and allowing it to vary is expected to have an effect on the model's forecasting performance.

Finally, in Figure 35, we observe that the measurement error standard deviations of our time varying COMPASS model are uniformly lower than the estimates of the standard fixed parameter model, implying that allowing for time variation in fact improves the fit of the model leading to smaller measurement error.

5.5 Time varying impulse response functions

In this Section, we investigate the changing transmission mechanism of shocks over time. Figure 36 and 37 display the impulse response functions for output, prices, nominal interest rate and the exchange rate to a monetary policy shock varying over time[52].

[51] In fact, Burgess et al. (2013) calibrate this parameter to a value of 1.007 (or 2.7902 in annualised terms).
[52] The monetary policy shock in the model is a transitory shock without persistence.

Figure 35: IRFs of variables to 1 st. dev. monetary policy shock

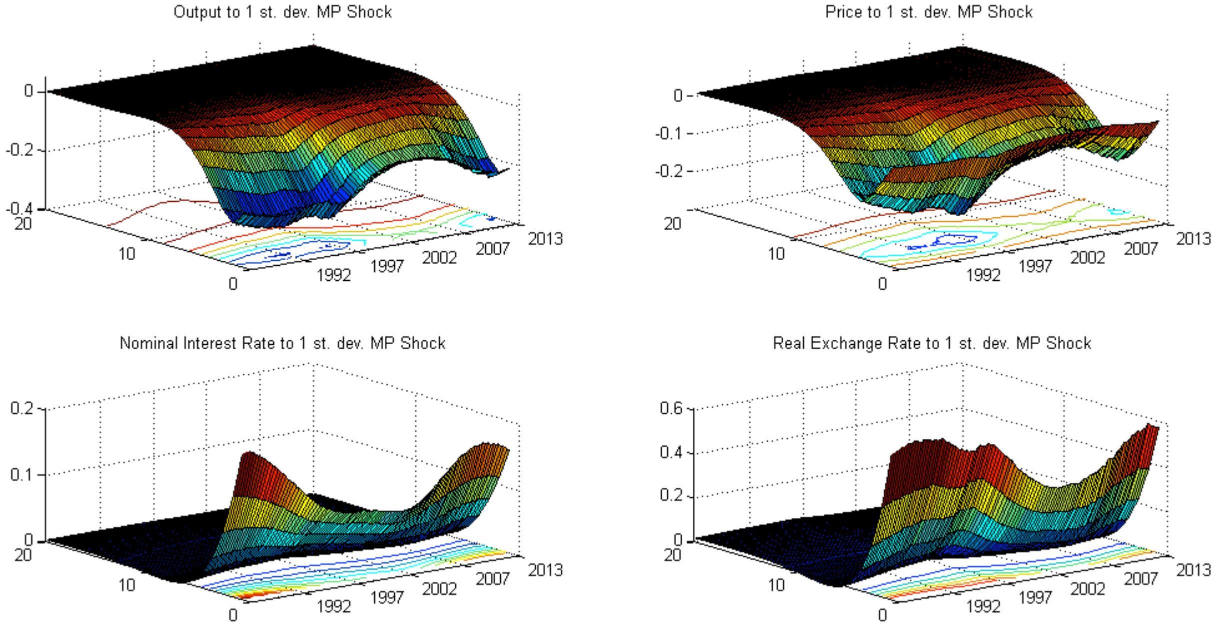


Figure 36 displays responses to a one standard deviation of the shock and captures the effect of the policy shock on the variables of interest, while also taking into account the changing size of the shock. Figure 37, on the other hand, presents the responses to a 25 basis points shock, which is useful in investigating changes in the transmission mechanism of the shock while keeping the size of the shock constant over time. It is evident from the estimation results in Figure 34 that the size of the monetary policy shock, that is, its volatility, displays a U-shape, with values of 0.1 in the period 1995-2005. A reduced role for monetary policy shocks over this period is consistent with policy having been very stable over this period. As result, both output and inflation become less responsive to the policy shock during this period, which is clear from Figure 36. From Figure 37, it is evident that even when we 'switch off' the changing volatility over time, by keeping the size of the shock fixed, output and inflation remain less responsive to the shock after 1997. These results are consistent with evidence presented in Boivin and Giannoni (2006) for the U.S., who interpret the decreased responsiveness of inflation and output to a monetary policy shock after the 1980s which they find in their application, as an outcome of the monetary authority becoming more effective and systematically more responsive in managing economic fluctuations after 1980. Similarly, for the UK, we can explain the decreased responsiveness of output and inflation to a policy shock (i.e. response to the non-systematic component of monetary policy) by the Bank of England becoming independent and hence more effective in the systematic conduct of monetary policy after 1998.

Figure 36: IRFs of variables to 25 basis points monetary policy shock

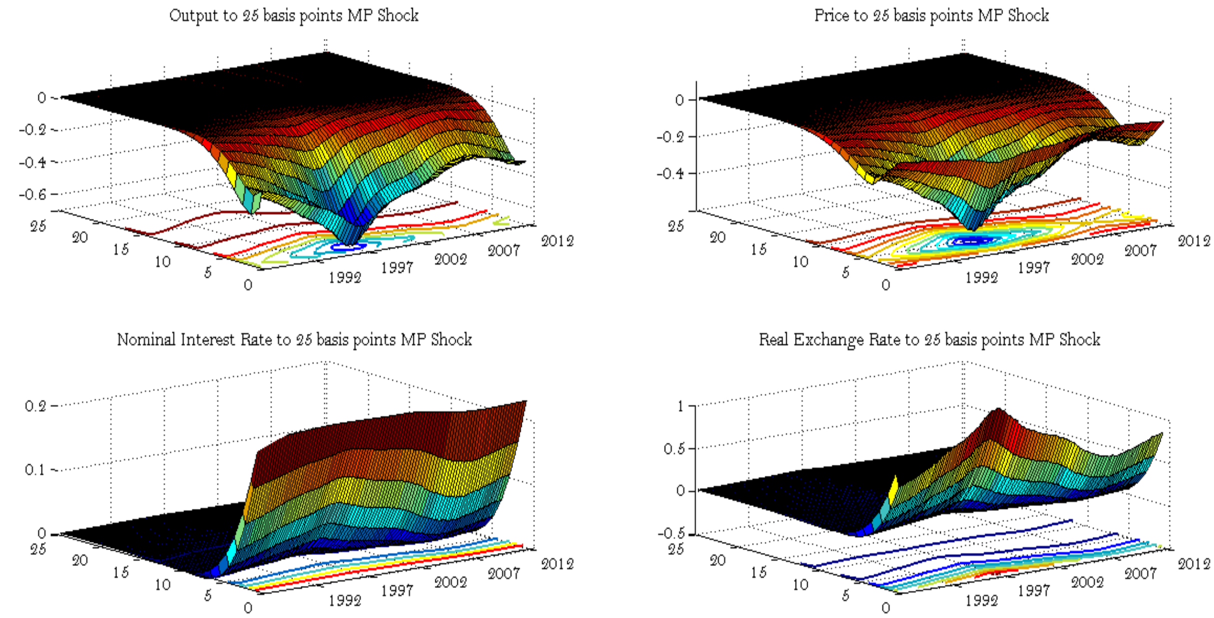
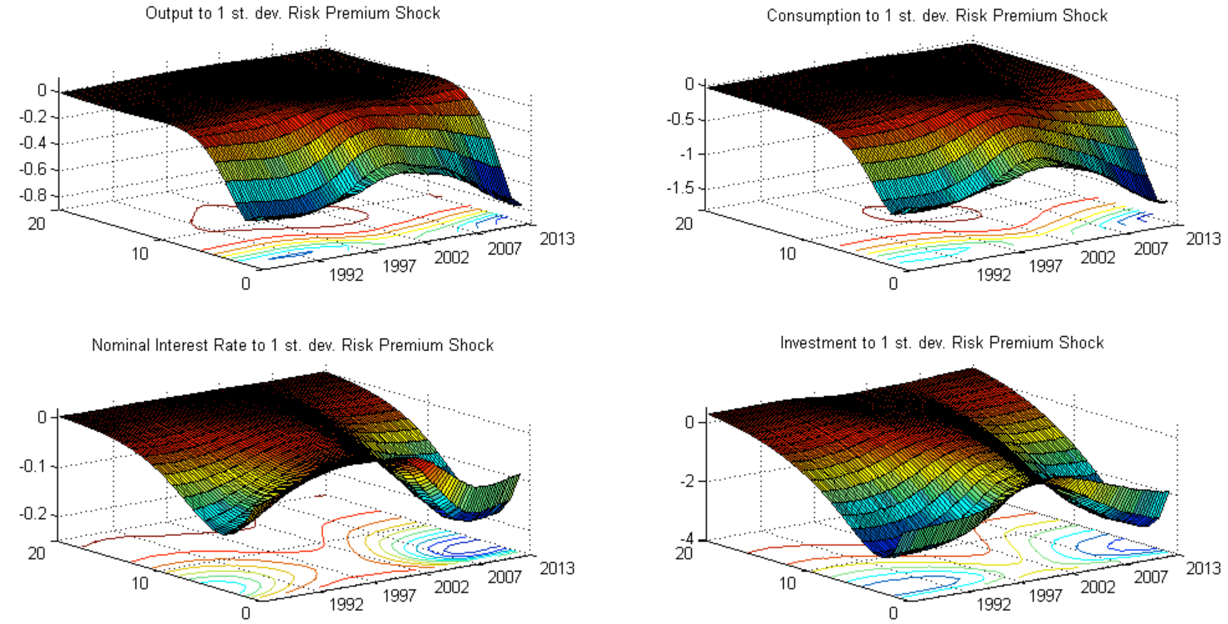


Figure 38 and 39 present the impulse response functions of output, consumption, the nominal rate and investment to: (i) a one standard deviation and, (ii) 25 basis points of the risk premium shock respectively. From the analysis in the previous Section, both the persistence and the volatility of the shock display a U-shape during 1995-2005, so it is not surprising that the responses to the shock follow a similar pattern with responses for output, consumption and investment all responding less to the shock both on impact and in terms of duration during this period. In addition, we find evidence of considerable time variation in the transmission mechanism of the risk premium shock to the nominal interest rate and this result is robust whether we consider the changing size of the shock, as in Figure 38, or 25 basis points, as in Figure 39. In particular, the response to a standard deviation of the shock is close to zero during the period of 1997-2005 on impact and on duration and jumps to -0.2 during the crisis period with around 15 quarters required for the nominal rate to go back to steady state. This large structural change in the nominal rate during the second half of the sample can be explained by the uncovered time variation in the Taylor rule output gap parameter during the same period, since the output gap is affected by the risk premium shock through its effect on consumption and investment. This, combined with the increase in the size of the risk premium shock, causes an intensified response of the interest rate.

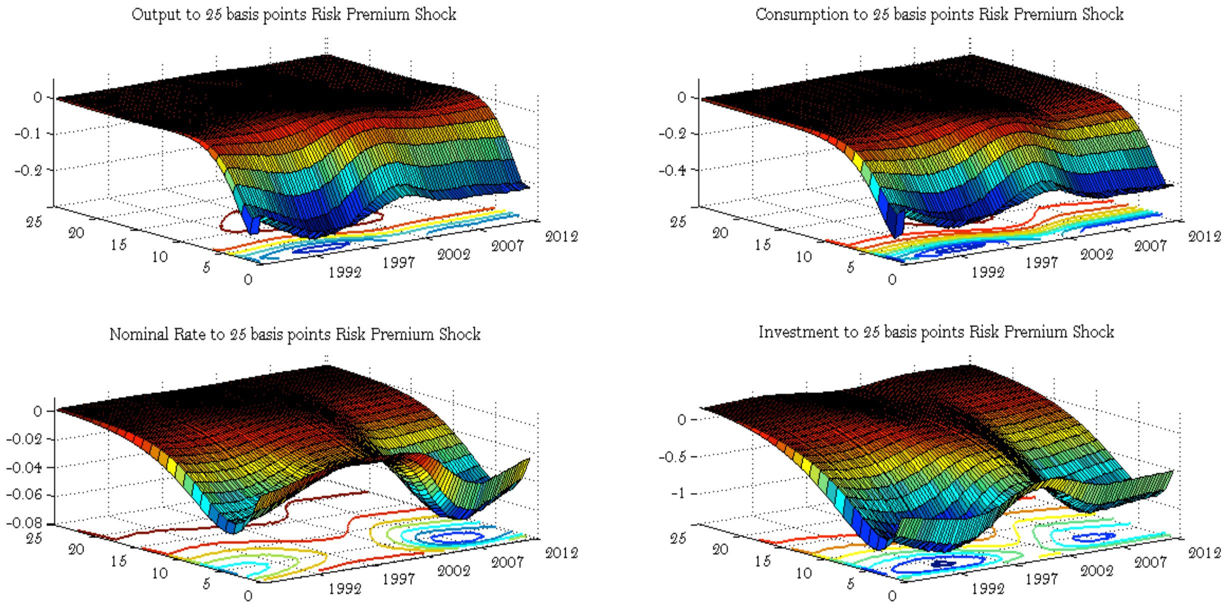
Figure 37: IRFs of variables to 1 st. dev. risk premium shock



5.6 Time varying variance decompositions

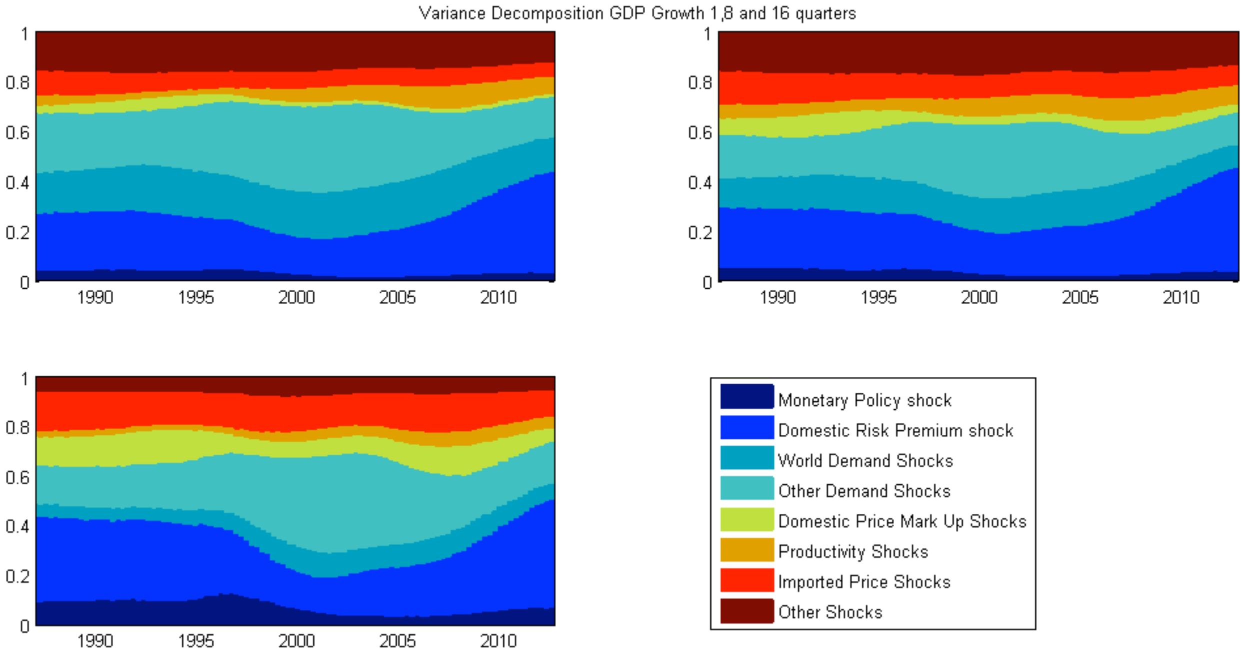
In this Section, we investigate the changing variance decompositions of key model observables over time. Figures 40-42 display the proportion of the variance of GDP growth, inflation and the policy rate respectively explained by the various exogenous shocks over time. The variance of GDP growth is explained primarily by the variance of demand shocks but towards the end of the sample, we see that the risk premium shock also starts to play a role even at longer horizons. Inflation's variation is absorbed almost entirely by the variance of domestic mark up shocks at one quarter ahead, while at two and four years, we observe that imported price shocks and the risk premium shock also have an effect. Finally, the variance of the policy rate is almost entirely explained by the variance of the monetary policy shock at one quarter ahead. In the period 1997-2005, domestic demand shocks explain around 20-40% of the policy rate variation at two and four years, while at the end of the sample, the risk premium shock explains nearly 80% of the policy rate variation. This is more evidence suggesting that the role of the risk premium shock has been substantial during the recent financial crisis. There are two competing explanations for these results. First, the risk premium is a wedge between the interest rate set by the central bank and the rates faced by agents in the model at which they borrow and lend and in this sense, the risk premium shock can be thought of as a shock to the borrowing costs of agents. So, unsurprisingly, its size increases during the

Figure 38: IRFs of variables to 25 basis points risk premium shock



recent financial crisis, reflecting the increasing borrowing costs during the so called credit crunch. A second and more critical approach to explaining the result is by recalling that COMPASS is a stylised DSGE model which lacks a financial sector and cannot therefore account for the events of 2008, other than through an increase in the variance of the exogenous risk premium shock.

Figure 39: Variance decomposition of output growth



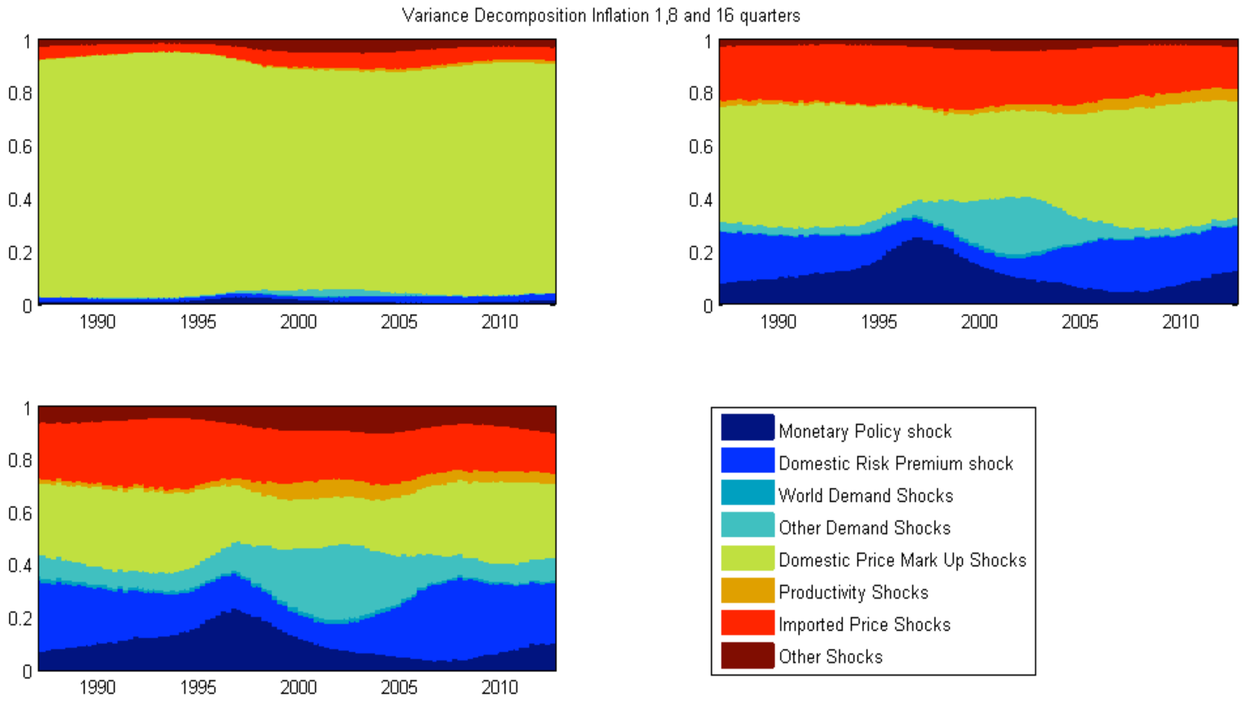
5.7 Forecasting

In this Section we evaluate the relative forecasting performance of our time varying parameter COMPASS (TVP-COMPASS) model. In addition, we compare the forecasting record of COMPASS against the fixed-parameter COMPASS (F-COMPASS) specification[53]. We measure accuracy of point forecasts using the root mean squared forecast error (RMSFE). The accuracy of density forecasts are measured by log predictive scores. We compute the logscore with the help of a nonparametric estimator to smooth the draws from the predictive density obtained for each forecast and horizon. We test whether the TVP-COMPASS model is statistically more accurate than the benchmark F-COMPASS with the Diebold and Mariano (1995) statistic computed with Newey-West estimator to obtain standard errors. We provide the results of the Diebold-Mariano two-sided test for the RMSFEs and logscores.

In addition, we also informally access the density forecast performance of the two models by looking at the probability integral transformation (PIT) at Figure 44, computed as the cumulative density function of the nonparametric estimator for the predictive density at the ex-post realised

[53]See Fawcett, Koerber, Masolo and Waldron (2015) for an evaluation of the forecast performance of a fixed parameter version of COMPASS against statistical and judgemental benchmarks.

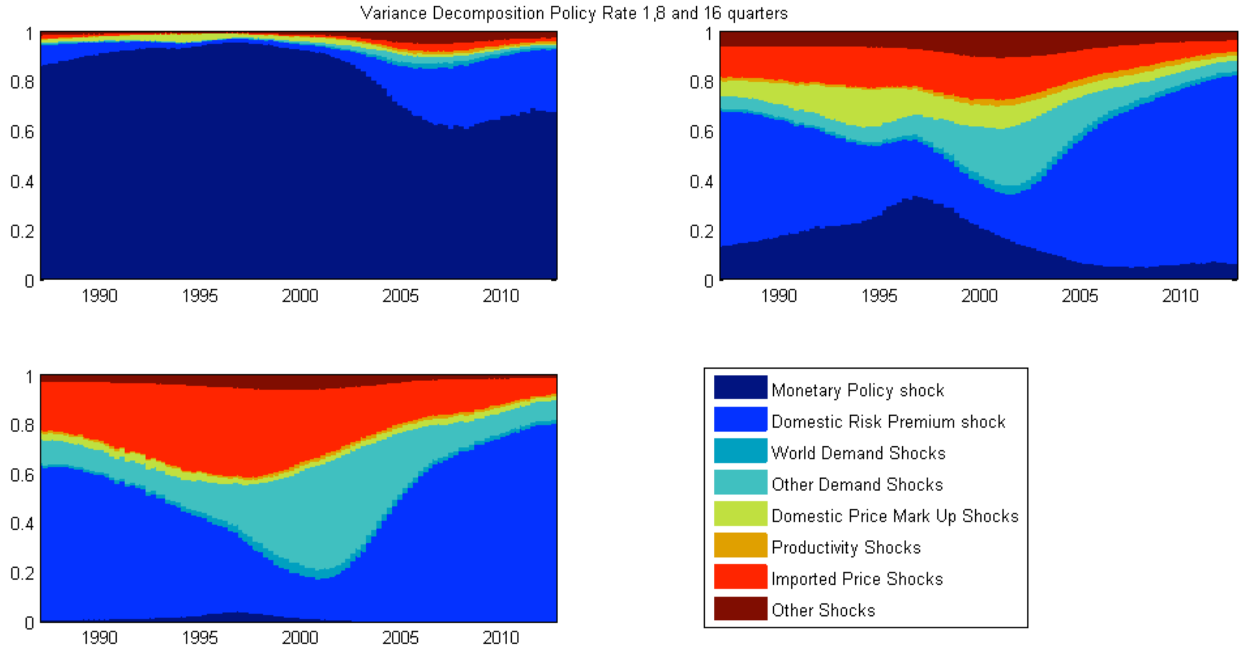
Figure 40: Variance decomposition of inflation



value of the target variable obtained for each forecast and horizon.

Table 19 presents the absolute performance of our TVP COMPASS model (in RMSFEs) and the relative performance of our approach to the standard F-COMPASS over different horizons (numbers smaller than one imply superior performance of the TVP COMPASS relative to the F-COMPASS). One, two and three stars indicate that we reject the null of equal accuracy in favour of the better performing model at significance levels of 10%, 5% and 1% respectively. In order to better understand the strengths and weaknesses of our time varying modification, we have split the forecast sample into two sub-periods: 1997Q3-2005Q2 and 2005Q3-2010Q4. Moreover, we estimate two different specifications of COMPASS: with and without estimating the productivity trend parameter. When this trend is not estimated, it is calibrated to a value of 1.007 as in Burgess et al. (2013). From Table 19, several conclusions emerge. First, in the relatively tranquil period of 1997Q3-2005Q2, which we argued in the previous Section is characterised by low volatility of monetary policy and risk premium shocks, we find better point forecasts for output growth, consumption growth and the interest rate. The better forecast performance of the TVP COMPASS model for the interest rate is due to the uncovered time variation in the policy rule parameters and in particular, the coefficient on output gap. Moreover, we find large significant improvements for

Figure 41: Variance decomposition of policy rate



inflation (17-19%) at short horizons, while the F-COMPASS performs better at longer horizons. Second, when looking at the period 2005Q3-2010Q4, we find statistically significant improvements for output, consumption and hours growth. On the other hand, our TVP COMPASS delivers slightly worse forecasts for inflation, but the differences are rarely statistically significant. Third, it seems that whether we estimate the productivity trend or keep it fixed makes little difference as far as forecasting is concerned. However, there is a pattern in the results; in the first subperiod, adding the trend to the parameter vector delivers almost always uniformly worse point forecasts (although the differences are very small). This is interesting and could be explained by the fact that the period is characterised by relatively stable productivity over time, hence as long as productivity is calibrated at a reasonable value, adding it to the parameter vector only reduces the degrees of freedom of the estimation, which makes forecasts worse. On the other hand, in the period 2005Q3-2010Q4, when trend productivity is falling, as shown in Figure 36, the forecasts generated with the trend specification are almost always uniformly better than when keeping trend calibrated.

RMSFEs Forecast Origins: 1997Q3: 2005Q2									RMSFEs Forecast Origins: 2005Q3:2010Q4						
TREND															
H		Y	C	I	INF	EXC	INT	H	Y	C	I	INF	EXC	INT	H
1		0.55	0.62	5.59	0.24	1.95	0.10	0.58	1.75	1.32	5.05	0.35	3.78	0.27	1.38
2	TV	0.51	0.58	4.82	0.21	1.72	0.17	0.60	1.42	1.55	4.40	0.40	3.55	0.51	1.07
4		0.49	0.61	5.35	0.23	1.60	0.23	0.60	1.27	1.52	5.00	0.41	3.44	0.91	0.81
8		0.46	0.56	4.76	0.27	1.57	0.19	0.54	1.22	1.32	4.86	0.37	3.29	1.41	0.78
1	TV	1.11	0.93	1.03**	0.81*	1.01	0.89**	1.08	0.96**	0.86**	1.00	1.04	1.01	0.99	0.99
2	/F	0.96	0.97	0.99	0.83	1.00	0.89	0.92**	0.89**	0.88**	0.96	1.09*	1.01	1.01	0.90**
4		0.89*	0.98	1.01	1.05	1.01	0.86	0.91*	0.93*	0.93*	1.02	1.09	1.00	1.03	0.92*
8		0.97	0.96	1.04*	1.08	0.99	0.93	1.00	0.97	0.97	1.02	1.05	1.00	0.99	1.00
NO TREND															
H		Y	C	I	INF	EXC	INT	H	Y	C	I	INF	EXC	INT	H
1		0.57	0.61	5.57	0.24	1.98	0.10	0.59	1.86	1.40	5.11	0.35	3.81	0.27	1.44
2	TV	0.49	0.56	4.76	0.21	1.74	0.18	0.59	1.52	1.62	4.62	0.39	3.58	0.51	1.13
4		0.47	0.60	5.28	0.23	1.60	0.24	0.60	1.32	1.56	5.30	0.41	3.45	0.93	0.83
8		0.45	0.56	4.74	0.27	1.58	0.20	0.54	1.25	1.35	5.10	0.37	3.29	1.46	0.78
1	TV	1.13	0.92	1.03**	0.81*	1.02	0.89**	1.11	1.02	0.91*	1.02	1.02	1.02*	1.00	1.02*
2	/F	0.98	0.97	0.99	0.83	1.00	0.88*	0.94*	0.96**	0.92**	1.02	1.08	1.02	1.03	0.94**
4		0.90	0.98	1.02	1.06	1.01	0.86	0.92*	0.98**	0.96	1.09	1.07	1.00	1.05	0.95**
8		0.97	0.96	1.05*	1.08	1.00	1.00	1.00	1.00	1.00	1.08	1.04	1.00	1.03	1.00

Table 19. RMSFEs. The figures under TV are absolute RMSFEs of the COMPASS estimated with QBLL, computed as the mean of the predictive density, the figures under the F are ratios of RMSFEs of TVP-COMPASS over the alternative fixed parameter COMPASS model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold - Mariano test.

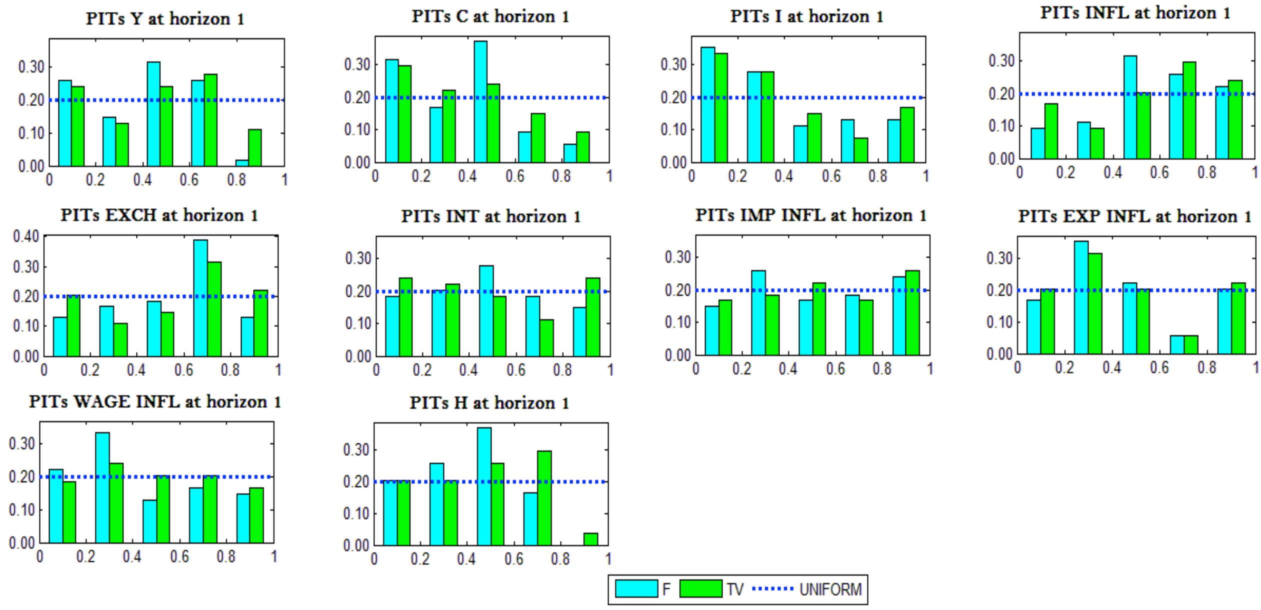
Table 20 accesses the quality of the density forecasts measured by logscores of the predictive density. The table displays absolute log predictive score for the TVP-COMPASS model and differences in logscores over the alternative F-COMPASS model, so numbers greater than zero imply superior performance of our approach. It is clear from Table 20 that allowing for time variation in the parameters of COMPASS delivers large and statistically significant improvements in the density forecasts during 1997Q3-2005Q2. The period of 2005Q3-2010Q4, on the other hand, is characterised by better TVP COMPASS performance for some variables (output, consumption and hours growth), while worse for others.

		Log Predictive Score: Forecast Origins: 1997Q3: 2005Q2							Log Predictive Score: Forecast Origins: 2005Q3:2010Q4						
		TREND													
H		Y	C	I	INF	EXC	INT	H	Y	C	I	INF	EXC	INT	H
1		-1.06	-1.02	-3.43	0.01	-2.11	0.85	-1.18	-2.27	-1.72	-3.15	-0.44	-2.88	-0.63	-1.73
2	TV	-1.08	-1.00	-2.99	0.06	-2.05	0.31	-1.23	-1.80	-1.93	-2.92	-0.62	-2.85	-1.99	-1.53
4		-1.10	-1.05	-3.17	-0.04	-2.03	0.05	-1.27	-1.66	-1.89	-3.09	-0.69	-2.91	-3.33	-1.43
8		-1.10	-1.02	-3.02	-0.15	-2.02	0.15	-1.26	-1.63	-1.78	-2.99	-0.55	-3.09	-6.44	-1.43
1		0.09**	0.20***	0.00	0.28***	0.10**	0.17*	0.12***	0.08**	0.22**	-0.02	-0.09	-0.02	-0.16	0.04
2	TV	0.13***	0.21***	0.04*	0.30***	0.12**	0.08	0.15***	0.19**	0.27**	0.03	-0.20	-0.10	-0.19	0.10***
4	-F	0.15**	0.21**	0.00	0.21**	0.13**	0.13*	0.15**	0.07**	0.11	-0.07	-0.22	-0.18	0.83	0.05*
8		0.14*	0.25**	-0.06*	0.14*	0.14*	0.18	0.14**	0.02	-0.02	0.00	-0.15	-0.43	1.41	0.03
		NO TREND													
H		Y	C	I	INF	EXCH	INT	H	Y	C	I	INF	EXCH	INT	H
1		-1.06	-1.02	-3.44	0.00	-2.12	0.85	-1.19	-2.39	-1.81	-3.19	-0.44	-2.87	-0.64	-1.78
2	TV	-1.07	-0.98	-2.98	0.06	-2.05	0.31	-1.23	-1.91	-2.02	-2.98	-0.61	-2.87	-2.00	-1.57
4		-1.09	-1.03	-3.15	-0.05	-2.03	0.00	-1.27	-1.72	-1.95	-3.17	-0.67	-2.98	-3.58	-1.44
8		-1.10	-1.02	-3.02	-0.15	-2.02	0.11	-1.26	-1.65	-1.81	-3.05	-0.56	-3.09	-6.51	-1.44
1		0.08***	0.20***	-0.03	0.28***	0.09**	0.18**	0.11***	-0.01	0.15	-0.06	-0.08	-0.01	-0.22	-0.01
2	TV	0.13***	0.22***	0.04	0.30***	0.12**	0.08	0.15***	0.08**	0.18*	-0.04	-0.19	-0.09	-0.23	0.06***
4	-F	0.15**	0.22**	0.00	0.20**	0.13**	0.10	0.15**	0.01	0.06	-0.15	-0.19	-0.25	0.58	0.05**
8		0.14*	0.24*	-0.06*	0.14*	0.15*	0.14	0.15**	-0.02	-0.07	-0.08	-0.15	-0.43	1.11	0.03

Table 20: Log Predictive Scores. The figures under TV are absolute log predictive scores for the COMPASS estimated with QBLL, computed as the log of the predictive density evaluated at the ex-post realised observation, the figures under F are differences of log scores of the TVP-COMPASS over the alternative fixed parameter model. '*', '**' and '***' indicate rejection of the null of equal performance against the two-sided alternative at 10%, 5% and 1% significance level respectively, using a Diebold-Mariano test.

Another way of assessing density forecast performance of the two models is by looking at the probability integral transformation (PITs) in Figure 44, computed as the CDF of the predictive density evaluated at the ex-post realised observation. Over a long enough sample, one would expect outturns in all parts of the distribution at frequencies that match the relevant probabilities. For example, we would expect to observe 5% of the realised values in the bottom 5% of the distribution. The histogram in Figure 44 displays the PITs for the TVP-COMPASS model and the F-COMPASS with green and blue bars respectively and the blue dotted line is the pdf of a uniform distribution. The PITs in Figure 44 for selected variables at one step ahead reveal that neither model is very close to delivering a uniform CDF. However, the TVP-COMPASS model is closer to being uniform than the F-COMPASS variant, suggesting that allowing for time variation improves forecast density accuracy at the one-step ahead horizon at least to some degree.

Figure 42: Probability Integral Transformations



5.8 Summary

This Chapter applied the quasi-Bayesian nonparametric procedure developed by Chapter 3 to a DSGE model of the UK economy, in order to assess the question of structural change in the model's parameters. We found evidence for time variation in the Taylor rule parameters as well as in the standard deviation and persistence parameters of some of the exogenous shocks. By looking at the impulse response functions and the variance decompositions, we also reached a conclusion about the importance of the risk premium shock and its role in the recent financial crisis. We offered two alternative explanations for this result. The first provides an account for the crisis through an increase in the borrowing costs of agents in the model, while the second explains the result through an increase in the volatility which shows up in the variances of the exogenous component of the model, by arguing that the DSGE model is too stylised to fit well the events of the crisis. Finally, this Chapter also provided a forecasting comparison of the time varying structural model with the standard fixed parameter model and we found gains for both point and density forecasts, especially in the 1997Q3-2005Q2 period.

6 Conclusion

This thesis establishes a novel quasi-Bayesian local likelihood (QBLL) approach for econometric inference in models with time varying parameters. The Bayesian framework is constructed by augmenting the local likelihood of Giraitis et al. (2016) through the Bayes principle with a prior distribution; this idea delivers asymptotically valid quasi-posterior distributions which admit closed form expressions in the special case of linear Gaussian models.

The approach is of sufficient generality and flexibility to produce Gibbs algorithms with mixtures of time varying and time invariant parameters. In addition, the approach can be applied to structural models where the quasi-posterior distributions are not of a known form, using the Metropolis algorithm proposed in Chapter 3.

Our Monte Carlo study indicates that the class of QBLL estimators exhibit good finite sample properties, and inference based on their quasi-posterior densities delivers valid confidence intervals. Crucially, as a consequence of modelling time variation nonparametrically, inference based on the QBLL approach is robust to different processes for the drifting parameters, its validity not depending on parametric restrictions typically imposed by state space models. In addition, the proposed approach has the capacity to facilitate a large number of variables, as a result of the Bayesian principle which allows to shrink parameters whenever the model is large. As demonstrated in Chapter 2, this can have important applications in forecasting, delivering significant improvements for both point and density forecasts.

In Chapter 2, we employ the novel approach to a VAR model to empirically address the issue of changing macroeconomic dynamics in the U.S. We uncover significant structural change in the series for core inflation, inflation persistence and the natural rate of unemployment as well as substantial drifts in the volatility of the series. We stress that in this context, not modelling explicitly the time variation will result in invalid inference on the model's parameters.

In Chapter 3 we developed the method further to the estimation of structural DSGE models and applied it to the Smets and Wouters (2007) model. In Chapter 4, we extend the DSGE model and add financial frictions in order to investigate the structural change in the financial sector of the model during the recent crisis. We found that the volatility of the financial shock of the model increases two-fold during the crisis period while the structural parameters guiding the financial sector remain unchanged, which led us to give a 'Bad Luck' event interpretation of the crisis. Finally, in Chapter 5, we estimated the COMPASS model used by the Bank of England on UK data in order to investigate structural change in the UK economy.

The main result that we find in Chapters 3-5 is that DSGE models' parameters are generally not constant over time. We find evidence of drifts in the policy parameters and this is unsurprising, given how much monetary policy has evolved in the past few decades. We also find considerable variation in the volatility of the structural shocks. Importantly, when allowing for parameter variation, we can improve the forecasting performance of standard fixed parameter models, which can provide a valuable tool for policy analysis.

References

Abramovich, Y., Spencer, N. and Turley, M. (2007). Time-varying autoregressive (TVAR) models for multiple radar observations, *IEEE Transactions on Signal Processing* **55**(4): 1298–1311.

Adolfson, M., Andersson, M., Linde, J., Villani, M. and Vredin, A. (2007). Modern forecasting models in action: Improving macroeconomic analyses at Central Banks, *International Journal of Central Banking* **3**(4): 111–144.

An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models, *Econometric Reviews* **26**(2-4): 113–172.

Bańbura, M., Giannone, D. and Reichlin, L. (2010). Large Bayesian vector autoregressions, *Journal of Applied Econometrics* **25**(1): 71–92.

Baumeister, C. and Benati, L. (2013). Unconventional monetary policy and the great recession: estimating the macroeconomic effects of spread compression at the zero lower bound, *International Journal of Central Banking* **9**(2): 165–212.

Benati, L. and Mumtaz, H. (2007). U.S. evolving macroeconomic dynamics: a structural investigation, *ECB: Working Paper Series* (746).

Benati, L. and Surico, P. (2009). VAR analysis and the Great Moderation, *American Economic Review* **99**(4): 1636–1652.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management, *Journal of Business and Economic Statistics* **19**(4): 465–474.

Bernanke, B., Gertler, M. and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework, *in* J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Vol. 1C, Elsevier, pp. 1341–1393.

Bernanke, B. and Mihov, I. (1998). Measuring monetary policy, *Quarterly Journal of Economics* **113**: 869–902.

Bianchi, F. (2013). Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics, *Review of Economic Studies* **80**: 463–490.

Bickel, P. J. and Yahav, J. A. (1969). Some contributions to the asymptotic theory of Bayes solutions, *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **11**(4): 257–276.

Blanchard, O. J. and Kahn, C. M. (1980). The solution of linear difference models under rational expectations, *Econometica* **48**(5): 1305–1312.

Boivin, J. and Giannoni, M. (2006). Has monetary policy become more effective?, *Review of Economics and Statistics* **88**(3): 445–462.

Born, B. and Pfeifer, J. (2011). Policy risk and the business cycle, *Bonn Econ Discussion Papers, University of Bonn, Germany* .

Burgess, S., Fernandez-Corugedo, E., Groth, C., Harrison, R., Monti, F., Theodoridis, K. and Waldron, M. (2013). The Bank of England's forecasting platform: COMPASS, MAPS, EASE and the suite of models, *Bank of England Working Paper Series, No471* .

Canova, F. (2006). Monetary policy and the evolution of the US economy, *CEPR Discussion Papers 5467* .

Canova, F. and Ferroni, F. (2012). The dynamics of US inflation: Can monetary policy explain the changes, *Journal of Econometrics* **167**(1): 47–60.

Canova, F. and Perez Forero, F. (2015). Estimating overidentified, nonrecursive, time-varying coeffcients structural vector autoregressions, *Quantitative Economics* **6**(2): 359–384.

Canova, F. and Sala, L. (2009). Back to square one: Identifcation issues in DSGE models, *Journal of Monetary Economics* **56**(4): 431–449.

Cantore, C., Levine, P. and Melina, G. (2012). Fiscal-monetary policy: A DSGE model fit for purpose, *Working Paper, University of Surrey* .

Cardani, R., Paccagnini, A. and Villa, S. (2015). Forecasting in a DSGE model with banking intermediation: evidence from the US, *Dems Working Paper Series* (292).

Carriero, A., Clark, T. and Marcellino, M. (2015a). Bayesian VARs: specification, choices and forecast accuracy, *Journal of Applied Econometrics* **30**: 46–73.

Carriero, A., Clark, T. and Marcellino, M. (2015b). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility, *Journal of the Royal Statistical Society (series A)* **178**(4): 837–862.

Carriero, A., Galvão, A. and Kapetanios (2015). A comprehensive evaluation of macroeconomic forecast methods, *Working Paper* .

Carriero, A., Kapetanios, G. and Marcellino, M. (2012). Forecasting government bond yields with large bayesian vector autoregressions, *Journal of Banking & Finance* **36**(7): 2026–2047.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrica* **81**: 541–553.

Castelnuovo, E. (2012). Estimating the evolution of money's role in the U.S. monetary business cycle, *Journal of Money, Credit and Banking* **44**(1): 23–52.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation, *Journal of Econometrics* **115**: 239–346.

Christiano, L., Eichenbaum, M. and Evans, C. (1999). Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Vol. 1A, Elsevier, pp. 65–148.

Christiano, L., Eichenbaum, M. and Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy, *Journal of Political Economy* **113**(1): 1–45.

Christiano, L. J., Motto, R. and Rostagno, M. (2014). Risk shocks, *American Economic Review* **104**(1): 27–65.

Christoffel, K., Coenen, G. and Warne, A. (2010). Forecasting with DSGE models, *in* M. Clements and D. Hendry (eds), *Oxford Handbook on Economic Forecasting*, Oxford University Press: Oxford, pp. 89–128.

Clarida, R., Gali, J. and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: Evidence and some theory, *Quarterly Journal of Economics* **115**(1): 147–180.

Clark, T. E. (2012). Real-time density forecasting from BVARs with stochastic volatility, *Journal of Business and Economic Statistics* **29**(3): 327–341.

Cogley, T., Primiceri, G. E. and Sargent, T. J. (2010). Inflation-gap persistence in the US, *American Economic Journal: Macroeconomics* **2(1)**: 43–69.

Cogley, T. and Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics, *in* B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 331–88.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post World War II US, *Review of Economic Dynamics* **8**: 262–302.

Cogley, T. and Sargent, T. J. (2009). Anticipated utility and rational expectations as approximations of Bayesian decision making, *International Economic Review* **49**: 185–221.

Crowder, M. (1988). Asymptotic expansions of posterior expectations, distributions, and densities for stochastic processes, *Annuals of the Institute for Statistical Mathematics* **40**: 297–309.

Del Negro, M. (2003). Discussion of Cogley and Sargent's 'drifts and volatilities: Monetary policy and outcomes of the post WWII U.S., *Federal Reserve of Atlanta Working Paper 26* .

Del Negro, M., Giannoni, M. and Schorfheide, F. (2014). Inflation in the Great Recession and new Keynesian models, *NBER Working Paper 20055* .

Del Negro, M., Hasegawa, R. and Schorfheide, F. (2014). Dynamic prediction pools: an inverstigation of financial frictions and forecasting performance, *Working Paper* .

Del Negro, M. and Schorfheide, F. (2013a). Bayesian econometrics, *in* J. Geweke, G. Koop and H. Van Dijk (eds), *The Oxford Handbook of Bayesian Econometrics*, Vol. 2, Oxford University Press: Oxford, pp. 293–390.

Del Negro, M. and Schorfheide, F. (2013b). DSGE model-based forecasting, *in* G. Elliott and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 2, Elsevier, pp. 57–140.

DeLong, J. B. (1997). Reducing inflation, *in* C. Romer and D. Romer (eds), *America's only peacetime inflation: the 1970's*, NBER Studies in Business Cycles.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business and Economic Statistics* **13**: 253–263.

Doan, T., Litterman, R. and Sims, C. (1984). Forecasting and Conditional Projection using Realistic Prior Distributions, *Econometric Reviews* **3**: 1–100.

Durbin, J. and Koopman, S. (2002). A simple and efficient simulation smoother for state space time series analysis, *Biometrika* **89**(3): 603–616.

Edge, R. and Guerkaynak, R. (2010). How useful are estimated DSGE model forecasts for central bankers, *Brookings Papers of Economic Activity* .

Edge, R., Guerkaynak, R. and Kisacikoglu, B. (2013). Judging the DSGE model by its forecasts, *Working Paper* .

Edge, R., Kiley, M. and Laforte, J. (2009). A comparison of forecast performance between Federal Reserve staff forecasts, simple reduced-form models, and a DSGE model, *Federal Reserve Board of Governors Finance and Economics, Discussion Paper Series* .

Fawcett, N., Koerber, L., Masolo, R. and Waldron, W. (2015). Evaluating UK point and density forecasts from an estimated DSGE model: the role of off-model information over the financial crisis, *Bank of England Working Paper Series, No538* .

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. (2007). Estimating macroeconomic models: a likelihood approach, *Review of Economic Studies* **74**: 1059–1087.

Fernandez-Villaverde, J. and Rubio-Ramirez, J. F. (2008). How structural are the structural parameters, *in* K. R. D. Acemoglu and M. Woodford (eds), *NBER Macroeconomics Annual 2007*, Vol. 22, Chicago: University of Chicago Press.

Foerster, A. T., Rubio-Ramirez, J. F., Waggoner, D. F. and Zha, T. A. (2014). Perturbation methods for Markov-switching DSGE models, *NBER Working Paper 258* .

Fuentes-Albero, C. (2014). Financial frictions, financial shocks, and aggregate volatility, *Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C.* .

Gali, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?, *American Economic Review* **89**(1): 249–271.

Gali, J. and Gambetti, L. (2009). On the sources of the Great Moderation, *American Economic Journal* **1**(1): 26–57.

Gali, J. and Gertler, M. (1999). Inflation dynamics: A structural econometric analysis, *Journal of Monetary Economics* **44**(2): 195–222.

Geweke, J. (1999). Computational experiments and reality, *University of Iowa,Manuscript* .

Geweke, J. (2005). *Contemporary Bayesian Econmetrics and Statistics*, Hoboken: John Wiley and Sons.

Giacomini, R. and Rossi, B. (2009). Model comparisons in unstable environments, *Cemmap Working Paper CWP13/12* .

Giannone, D., Lenza, M. and Primiceri, G. E. (2015). Prior selection for vector autoregressions, *Review of Economics and Statistics* **97**(2): 436–451.

Gilchrist, S. and Zakrajsek, E. (2011). Credit spreads and business cycle fluctuations, *NBER Working Paper 17021* .

Giraitis, L., Kapetanios, G. and Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change, *Journal of Econometrics* **177**: 153–170.

Giraitis, L., Kapetanios, G., Theodoridis, K. and Yates, T. (2013). Estimating time-varying DSGE models using minimum distance methods, *Working Paper* .

Giraitis, L., Kapetanios, G., Wetherilt, A. and Zikes, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market, *Journal of Applied Econometrics* **31**(1): 58–84.

Giraitis, L., Kapetanios, G. and Yates, T. (2014). Inference on stochastic time-varying coefficient models, *Journal of Econometrics* **179**(1): 46–65.

Goodfriend, M. and King, R. (1997). The new neoclassical synthesis and the role of monetary policy, *in* B. Bernanke and J. Rotemberg (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 231–83.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chain and their applications, *Biometrika* **57**: 97–109.

Ibragimov, I. and Has'minskii, R. (1981). *Statistical Estimation: Asymptotic Theory*, Springer.

Ireland, P. (2004). Technology shocks in the new Keynesian model, *The Review of Economics and Statistics* **86**(4): 923–936.

Jacquier, E., Polson, N. G. and Rossi, P. (1994). Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics* **12**: 371–418.

Johannes, M., Lochstoer, L. and Mou, Y. (2015). Learning about consumption dynamics, *Journal of Financial Innovation* **1**(1): Forthcoming.

Justiniano, A. and Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations, *American Economic Review* **98**(3): 604–641.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR models, *Journal of Applied Econometrics* **12**(2): 99–132.

Kapetanios, G., Marcellino, M. and Venditti, F. (2015). Large time-varying parameter var: kernel-based mixed estimators, *Working Paper* .

Kim, C. and Nelson, C. R. (1999). Has the U.S. Economy become more stable? A Bayesian Approach based on a Markov-Switching Model of the Business Cycle, *Review of Economics and Statistics* **81**(4): 608–618.

Kim, J. Y. (1998). Large sample properties of posterior densities, Bayesian information criterion, and the likelihood principle in non-stationary time series models, *Econometrica* **62**(2): 359–380.

Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**: 361–393.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**(1): 1 – 25.

Kolasa, M. and Rubaszek, M. (2015). Forecasting with DSGE models with financial frictions, *International Journal of Forecasting* **31**: 1–19.

Koop, G. (2011). Forecasting with medium and large Bayesian VARs, *Journal of Applied Econometrics* .

Koop, G. and Korobilis, D. (2013). Large time-varying parameter vars, *Journal of Econometrics* **177**: 185–198.

Kreps, D. (1998). Anticipated utility and dynamic choice, *Frontiers of Research in Economic Theory, Cambridge University Press* pp. 242–274.

Kydland, F. and Prescott, E. (1996). The computational experiment: an econometric tool, *Journal of Economic Perspectives* **10**: 69–85.

Lindley, D. (1971). *Bayesian Statistics: A Review*, Philadelphia, SIAM.

Litterman, R. (1980). A Bayesian Procedure for Forecasting with Vector Autoregression, *Working Paper, Massachusetts Institute of Technology, Department of Economics* .

McConnell, M. and Perez Quiros, G. (2000). Output fluctuations in the U.S.: what has changed since the early 1980s?, *American Economic Review* **90**: 1464–1476.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21**(6): 1087–1092.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Mathematical Statistics.

Mumtaz, H. and Surico, P. (2009). Time-varying yield curve dynamics and monetary policy, *Journal of Applied Econometrics* **24**(6): 895–913.

Philippon, T. (2009). The bond market's Q, *Quarterly Journal of Economics* **124**: 1011–1056.

Phillips, P. (1995). Automated forecasts of Asia-Pacific economic activity, *Technical report*, Cowles Foundation Discusssion Papers.

Primiceri, G. (2005). Time-varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**(3): 821–852.

Roberts, G., Gelman, A. and Gilks, W. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms, *The Annals of Applied Probability* **7**(1): 110–120.

Rotemberg, J. and Woodford, M. (1997). An optimization-based econometric framework for the evaluation of monetary policy, *in* B. S. Bernanke and K. Rogoff (eds), *NBER Macroeconomics Annual*, MIT Press: Cambridge, pp. 297–346.

Rubaszek, M. and Skrzypczynski, P. (2008). On the forecasting performance of a small-scale DSGE model, *International Journal of Forecasting* **24**(3): 498–512.

Schorfheide, F. (2000). Loss function-based evaluation of DSGE model, *Journal of Applied Econometrics* **15**(6): 645–670.

Schorfheide, F. (2005). Learning and monetary policy shifts, *Review of Economic Dynamics* **8**(2): 392–419.

Schorfheide, F. (2007). Comment on 'how structural are structural parameters?', *Working Paper* .

Sims, C. (2002). Solving linear rational expectations models, *Computational Economics* **20**(1-2): 1–20.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

Sims, C. A. (2001). Comment on Sargent and Cogley's 'evolving post WWII US inflation dynamics, *NBER Macroeconomics Annual* **16**: 373–379.

Sims, C. A. and Zha, T. (2006). Were there regime switches in U.S. monetary policy?, *American Economic Review* **96**: 1193–1224.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the Euro Area, *Journal of the European Economic Association* **1**(5): 1123–1175.

Smets, F. and Wouters, R. (2005). Comparing shocks and frictions in U.S. and Euro Area business cycles: a Bayesian DSGE approach, *Journal of Applied Econometrics* **20**(2): 161–83.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach, *American Economic Review* **97**(3): 586–606.

Steinbach, R., Mathuloe, P. and Smith, B. (2009). An open economy new Keynesian DSGE model of the South African economy, *South African Journal of Economics* **77**: 207–227.

Stock, J. (2001). Discussion of Sargent and Cogley's 'evolving post WWII US inflation dynamics, *NBER Macroeconomics Annual* **16**: 379–387.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics* **14**: 11–30.

Stock, J. H. and Watson, M. W. (2003). How did leading indicator forecasts perform during the 2001 recession?, *Federal Reserve Bank of Richmond, Economic Quarterly* **89**(3): 71–90.

Taylor, J. B. (1993). Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy* **39**: 195–214.

Tian, L., Liu, J. S. and Wei, L. (2007). Implementation of estimating function-based inference procedures with Markov chain Monte Carlo samplers, *Journal of American Statistical Associations* **102**(479): 881–888.

Walker, A. M. (1969). On the asymptotic behavior of posterior distributions, *Journal of Royal Statistical Society* **31**(1): 80–88.

Wieland, V. and Wolters, M. (2011). The diversity of forecasts from macroeconomic models of the US economy, *Economic Theory* **47**: 247–292.

Woodford, M. (2003). *Interest and Prices: Foundation of a Theory of Monetary Policy*, Princeton University Press: Princeton.

Woulters, M. H. (2012). Evaluating point and density forecast of DSGE models, *Munich Personal RePEc Archive Paper n. 36147* .

7 Appendix

7.1 Proofs and Additional Results for Chapter 2

7.1.1 Proof of Proposition 1

We follow the proof of Theorem 1 in Chernozhukov and Hong (2003), which is a generalisation of earlier results by Ibragimov and Has'minskii (1981) and Bickel and Yahav (1969) for a non-likelihood based objective functions. We replace Chernozhukov and Hong (2003)'s asymptotic analysis based on $\sqrt{T}$-neighbourhoods of the true (in their analysis fixed) parameter θ^0 to general $\varkappa_{jT}^{1/2}$-neighbourhoods of θ_j^0 (now indexed by time) with $\varkappa_{jT}$ defined in (6). This is justified since

$$\varkappa_{jT} \sim H \frac{\left(\int K(x)dx\right)^2}{\int K^2(x)dx}$$

where H, the bandwidth parameter associated with the kernel, satisfies $H \to \infty$ as $T \to \infty$.

In fact, the argument of Chernozhukov and Hong (2003) is valid for an arbitrary $m_T^{1/2}$-neighbourhood as long as $m_T \to \infty$ as $T \to \infty$. The radius of the specified neighbourhood corresponds to the consistency rate of the extremum estimator of θ_j^0.

Assumptions.

For all $j = [\tau T]$, $0 < \tau < 1$:

1. $\theta_j^0 \in int(\Theta)$ and $\Theta \subset \mathbb{R}^{\dim(\theta_j)}$ is compact.

2. $\forall \delta > 0$, $\exists \lambda > 0$ such that

$$\liminf_{T\to\infty} P\left\{ \sup_{\|\theta_j - \theta_j^0\| > \delta} \frac{1}{\varkappa_{jT}} \left(\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0)\right) \le -\lambda \right\} = 1.$$

3. For θ_j in an open neighbourhood of θ_j^0, $\varphi_{Tj}(.)$ admits the following expansion

$$\varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) = (\theta_j - \theta_j^0)'\nabla\varphi_{Tj}(\theta_j^0) - \frac{1}{2}(\theta_j - \theta_j^0)'\varkappa_{jT} J_{Tj}(\theta_j^0)(\theta_j - \theta_j^0) + R_{Tj}(\theta_j) \quad (46)$$

where the main components

$$\nabla\varphi_{Tj}(\theta_j^0) = \frac{\partial\varphi_{Tj}(\theta_j^0)}{\partial\theta_j} \quad \text{and} \quad J_{Tj}(\theta_j^0) := -\frac{1}{\varkappa_{jT}} \mathbb{E}\mathcal{H}\left(\varphi_{Tj}\left(\theta_j^0\right)\right),$$

and the remainder $R_{Tj}(.)$ satisfy:

(a) $-J_{Tj}(\theta_j^0)^{1/2}\nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{jT}} \to_d \mathcal{N}(0, I)$ as $T \to \infty$.

(b) The matrix $J_{Tj}(\theta_j^0)$ satisfies

$$\begin{aligned} \lambda^* &:= \limsup_{T\to\infty} \lambda_{\max}(J_{Tj}(\theta_j^0)) < \infty \\ \lambda_* &:= \liminf_{T\to\infty} \lambda_{\min}(J_{Tj}(\theta_j^0)) > 0, \end{aligned} \tag{47}$$

where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximal and minimal eigenvalue of a symmetric matrix A.

(c) $\forall\varepsilon > 0$, $\forall\eta > 0$, $\exists\delta > 0$ $\exists M > 0$ such that

$$\limsup_{T\to\infty} P\left\{ \sup_{M/\sqrt{\varkappa_{jT}}\le\|\theta-\theta_j^0\|\le\delta} \frac{\|R_{Tj}(\theta)\|}{\varkappa_{jT}\left\|\theta-\theta_j^0\right\|^2} > \eta \right\} \le \varepsilon, \tag{48}$$

$$\limsup_{T\to\infty} P\left\{ \sup_{\|\theta-\theta_j^0\|\le M/\sqrt{\varkappa_{jT}}} \|R_{Tj}(\theta)\| > \varepsilon \right\} = 0. \tag{49}$$

4. The prior density $\pi_j(.)$ is strictly positive and Lipschitz continuous function over Θ.

5. The time variation in the true parameters θ_j^0 satisfies one of conditions (1) or (2).

The asymptotics in Proposition 1 is established as $T \to \infty$, and since $j = [\tau T]$, we have $j \to \infty$. We explicitly allow the quantities above to depend on j in order to emphasise the time varying nature of the estimation problem. Assumption 1 is a standard assumption maintained by Giraitis et al. (2016). Assumption 2 is a high-level assumption that requires uniform convergence of the objective function outside a neighbourhood of θ_j^0. A CLT is assumed in 3a) for $\nabla\varphi_{Tj}(\theta_j^0)$, the gradient of the objective function. Assumption 3c), requires that the remainder term $R_{Tj}(\theta_j)$ of the quadratic approximation of the objective function is well behaved. By Lemma 2 in Chernozhukov and Hong (2003) Assumption 3c) is satisfied for any twice continuously differentiable objective function with a uniform convergence of the Hessian of second derivatives to its limit. Note that the objective functions studied in Giraitis et al. (2016) and in this thesis are twice continuously differentiable and in addition, Giraitis et al. (2016) assume uniform law of large numbers (ULLN) for the Hessian. Assumption 4 explicitly allows the prior density to vary over time and requires stronger Lipschitz continuity of π_j. Finally, Assumption 5 is required for consistent estimation in the frequentist setup of Giraitis et al. (2016) and is a necessary condition for the CLT in Assumption 3a).

Proof. We follow the Chernozhukov and Hong (2003) notation and define $h_{Tj} = \sqrt{\varkappa_{jT}}(\theta_j - \xi_{Tj})$ and $\xi_{Tj} := \theta_j^0 + \frac{1}{\varkappa_{jT}} J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)$. Letting

$$\mathbb{H}_{Tj} := \left\{ h_{Tj} \equiv \sqrt{\varkappa_{jT}} \left(\theta_j - \theta_j^0\right) - J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{jT}} : \theta_j \in \Theta \right\} \tag{50}$$

we need to show that

$$\int_{\mathbb{H}_{Tj}} \|h\|^a \left| p_{Tj}^*(h) - p_\infty(h) \right| dh \to_p 0 \tag{51}$$

for all $a \geq 0$. Recalling the definition of the quasi-posterior density in (7):

$$p_{Tj}(\theta_j) = \frac{\pi_j(\theta_j)\exp(\varphi_{Tj}(\theta_j))}{\int_\Theta \pi_j(\theta)\exp(\varphi_{Tj}(\theta))d\theta},$$

we begin by applying the transformation of $\theta_j = h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj}$ to obtain

$$\begin{aligned} p_{Tj}(h_{Tj}) &= \frac{1}{\sqrt{\varkappa_{jT}}} p_{Tj}(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj}) \\ &= \frac{\pi_j(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj})\exp(\varphi_{Tj}(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj}))}{\int_{\mathbb{H}_{Tj}} \pi_j(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj})\exp(\varphi_{Tj}(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj})dh} \qquad (52) \\ &= \pi_j(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj})\exp(\omega_{Tj}(h_{Tj}))/C_{Tj} \qquad (53) \end{aligned}$$

where

$$\omega_{Tj}(h_{Tj}) = \varphi_{Tj}\left(\frac{h_{Tj}}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}\right) - \varphi_{Tj}(\theta_j^0) - \frac{1}{2\varkappa_{jT}} \left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1} \left[\nabla\varphi_{Tj}(\theta_j^0)\right] \tag{54}$$

and

$$C_{Tj} = \int_{\mathbb{H}_{Tj}} \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) \exp\left\{\omega_{Tj}(h)\right\} dh. \tag{55}$$

For h_{Tj} belonging to the integration area $\mathbb{H}_{Tj}$ in (50), the following useful identity applies:

$$\omega_{Tj}(h_{Tj}) = -\frac{1}{2} h_{Tj}' J_{Tj}(\theta_j^0) h_{Tj} + R_{Tj}\left(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj}\right) \quad \forall h_{Tj} \in \mathbb{H}_{Tj}. \tag{56}$$

To prove (56), note that, by definition of $\mathbb{H}_{Tj}$, for any $h_{Tj} \in \mathbb{H}_{Tj}$ there exists a $\theta_j \in \Theta$ satisfying the identities

$$\begin{aligned} h_{Tj} &= \sqrt{\varkappa_{jT}}\left(\theta_j - \theta_j^0\right) - J_{Tj}(\theta_j^0)^{-1}\nabla\varphi_{Tj}(\theta_j^0)/\sqrt{\varkappa_{jT}}, \quad (57) \\ \theta_j &= \frac{h_{Tj}}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}. \quad (58) \end{aligned}$$

Applying (57) to (54) and using (46) of Assumption 3 we obtain,

$$\begin{aligned} \omega_{Tj}(h_{Tj}) &= \varphi_{Tj}(\theta_j) - \varphi_{Tj}(\theta_j^0) - \frac{1}{2\varkappa_{jT}}\left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1}\left[\nabla\varphi_{Tj}(\theta_j^0)\right] \\ &= (\theta_j - \theta_j^0)'\nabla\varphi_{Tj}(\theta_j^0) - \frac{1}{2}(\theta_j - \theta_j^0)'\varkappa_{jT}J_{Tj}(\theta_j^0)(\theta_j - \theta_j^0) + R_{Tj}(\theta_j) \\ &\quad - \frac{1}{2\varkappa_{jT}}\left[\nabla\varphi_{Tj}(\theta_j^0)\right]' J_{Tj}(\theta_j^0)^{-1}\left[\nabla\varphi_{Tj}(\theta_j^0)\right]. \end{aligned}$$

Using (58) in each of the terms in of the above expression and collecting terms proves (56). Having established (56), we prove that all subsequential probability limits of C_{Tj} in (55) are strictly positive. Using (49), dominated convergence and the properties of the Gaussian density we obtain

$$\begin{aligned} C_{Tj} &= \int_{\mathbb{H}_{Tj}} \pi_j\left(\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}\right)\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h + R_{Tj}\left(\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}\right)\right\}dh \\ &= \pi_j\left(\theta_j^0\right)\int_{\mathbb{R}^{\dim(\theta_j^0)}}\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h\right\}dh \\ &\quad + \int_{\mathbb{H}_{Tj}}\left\{\pi_j\left(\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}\right) - \pi_j\left(\theta_j^0\right)\right\}\exp\left\{-\frac{1}{2}h'J_{Tj}(\theta_j^0)h\right\}dh + o_p(1) \\ &= (2\pi)^{\dim(\theta_j^0)/2}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right) + o_p(1), \end{aligned}$$

because Lipschitz continuity of π_j implies that

$$\left\|\pi_j\left(\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}\right) - \pi_j\left(\theta_j^0\right)\right\| \leq c\left\|\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj} - \theta_j^0\right\| \to_p 0 \quad (59)$$

for all $h_{Tj} \in \mathbb{H}_{Tj}$. We conclude that, as $T \to \infty$

$$\left|C_{Tj} - (2\pi)^{\dim(\theta_j^0)/2}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right)\right| \to_p 0 \quad (60)$$

and, since $\liminf_{T\to\infty}\det(J_{Tj}(\theta_j^0))^{-1/2}\pi_j\left(\theta_j^0\right) > 0$ by Assumptions 3b and 4,

$$\liminf_{T\to\infty} P\left(C_{Tj} > \varepsilon\right) \geq 1 - \varepsilon \qquad (\forall\varepsilon > 0). \quad (61)$$

Using (53), the left side of (51) can be written as

$$\int_{\mathbb{H}_{Tj}} \|h\|^a \left| p^*_{Tj}(h) - p_\infty(h) \right| dh = A_{Tj} C_{Tj}^{-1} \tag{62}$$

where C_{Tj} is defined in (55) and

$$A_{Tj} = \int_{\mathbb{H}_{Tj}} \|h\|^a \left| \pi(\frac{h}{\sqrt{\varkappa_{jT}}} + \xi_{Tj}) \exp(\omega(h)) - C_{Tj}\phi_{Tj}(h) \right| dh$$

where

$$\phi_{Tj}(h_{Tj}) = (2\pi)^{-\dim(\theta_j)/2} \det(J_{Tj}(\theta_j^0))^{1/2} \exp\left\{ -\frac{1}{2} h'_{Tj} J_{Tj}(\theta_j^0) h_{Tj} \right\}.$$

Adding and subtracting $\exp\left\{-\frac{1}{2}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}\right\}\pi_j\left(\theta_j^0\right)$ in the above expression we obtain that $A_{Tj} \leq A_{Tj}^{(1)} + A_{Tj}^{(2)}$ where

$$\begin{aligned} A_{Tj}^{(1)} &= \int_{\mathbb{H}_{Tj}} \|h\|^a \left| \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) e^{\omega(h)} - e^{-\frac{1}{2}h' J_{Tj}(\theta_j^0) h} \pi_j\left(\theta_j^0\right) \right| dh, \\ A_{Tj}^{(2)} &= \int_{\mathbb{H}_{Tj}} \|h\|^a \left| \pi_j\left(\theta_j^0\right) - C_{Tj}(2\pi)^{-\dim(\theta_j)/2} \det(J_{Tj}(\theta_j^0))^{1/2} \right| \exp\left\{ -\frac{1}{2} h' J_{Tj}(\theta_j^0) h \right\} dh. \end{aligned} \tag{63}$$

By using (60) and the dominated convergence theorem, $A_{Tj}^{(2)} \to_p 0$ as $T \to \infty$. Also, (61) implies that $\left|C_{Tj}^{-1}\right| = O_p(1)$. Therefore, by (62), the proposition is proved if $A_{Tj}^{(1)} \to_p 0$ where $A_{Tj}^{(1)}$ the integral in (63).

To show that the integral $A_{Tj}^{(1)} \to_p 0$, defined in (63), we partition the area of integration in (50) as follows: $\mathbb{H}_{Tj} = \mathbb{H}_{Tj}^{(1)} \cup \mathbb{H}_{Tj}^{(2)} \cup \mathbb{H}_{Tj}^{(3)}$, where

$$\begin{aligned} \text{i) } \mathbb{H}_{Tj}^{(1)} &= \{h \in \mathbb{H}_{Tj} : \|h\| \leq M\}, \\ \text{ii) } \mathbb{H}_{Tj}^{(2)} &= \left\{h \in \mathbb{H}_{Tj} : \|h\| > \delta\sqrt{\varkappa_{jT}}\right\}, \\ \text{iii) } \mathbb{H}_{Tj}^{(3)} &= \left\{h \in \mathbb{H}_{Tj} : M < \|h\| \leq \delta\sqrt{\varkappa_{jT}}\right\}, \end{aligned}$$

where δ is some positive number satisfying (48) and M is a fixed number in $(0, \infty)$. It remains to prove $A_{Tj}^{(1)} \to_p 0$ over areas i)-iii). Note that, by (50) and Assumption 3(a),

$$\frac{h_{Tj}}{\sqrt{\varkappa_{jT}}} = \theta_j - \theta_j^0 + O_p\left(\frac{1}{\sqrt{\varkappa_{jT}}}\right)$$

so the areas i)-iii) can be expressed in terms of θ_j: for example,

$$\mathbb{H}^{(3)}_{Tj} = \left\{ \theta \in \Theta : \frac{M}{\sqrt{\varkappa_{jT}}} < \|\theta\| \leq \delta \right\}. \tag{64}$$

for all but finitely many T.

(i) Area $\mathbb{H}^{(1)}_{Tj}$. By finiteness of M, it suffices to show that the following quantity is $o_p(1)$:

$$\begin{aligned}
& \sup_{h\in\mathbb{H}^{(1)}_{Tj}} \|h\|^a \left| e^{\omega_{Tj}(h)} \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) - e^{-\frac{1}{2}h' J_{Tj}(\theta_j^0)h} \pi_j(\theta_j^0) \right| \\
= & \max_{\|h\|\leq M} \|h\|^a e^{-\frac{1}{2}h' J_{Tj}(\theta_j^0)h} \left| \exp\left\{ R_{Tj}\left(h/\sqrt{\varkappa_{jT}} + \xi_{Tj} \right) \right\} \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
\leq & M^a \max_{\|h\|\leq M} \left| \exp\left\{ R_{Tj}\left(h/\sqrt{\varkappa_{jT}} + \xi_{Tj} \right) \right\} \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
\leq & M^a \max_{\|h\|\leq M} \exp\left\{ R_{Tj}\left(h/\sqrt{\varkappa_{jT}} + \xi_{Tj} \right) \right\} \max_{\|h\|\leq M} \left| \pi_j(h/\sqrt{\varkappa_{jT}} + \xi_{Tj}) - \pi_j(\theta_j^0) \right| \\
& + M^a \max_{\|h\|\leq M} \left| \exp\left\{ R_{Tj}\left(h/\sqrt{\varkappa_{jT}} + \xi_{Tj} \right) \right\} - 1 \right| \sup_j \pi_j(\theta_j^0)
\end{aligned} \tag{65}$$

The second term of (65) is $o_p(1)$ by (49) since $\sup_{\|h\|\leq M} \left\| h/\sqrt{\varkappa_{jT}} + \xi_T - \theta_j^0 \right\| = O_p(1/\sqrt{\varkappa_{jT}})$. Since $\exp\left\{ R_{Tj}\left(h/\sqrt{\varkappa_{jT}} + \xi_{Tj} \right) \right\} = 1 + o_p(1)$ by (49), the first term of (65) is is $o_p(1)$ by (59).

(ii) Area $\mathbb{H}^{(2)}_{Tj}$. It is sufficient to show that:

a) $\int_{\|h\|>\delta\sqrt{\varkappa_{jT}}} \|h\|^a \pi(h/\sqrt{\varkappa_{jT}} + \xi_T) \exp(\omega(h)) dh \to_p 0$ and

b) $\int_{\|h\|>\delta\sqrt{\varkappa_{jT}}} \|h\|^a \exp\left\{ -\frac{1}{2}h' J_{Tj}(\theta_j^0)h \right\} dh \to_p 0$.

To show a), change variables $\theta_j = h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj}$, so that the area of integration is $\left\| \theta - \xi_{Tj} \right\| > \delta$. We have that the expression in a) is bounded by

$$\begin{aligned}
\varkappa_{jT}^{(\alpha+1)/2} \int_{\left\|\theta-\xi_{Tj}\right\|>\delta} & \left\| \theta - \xi_{Tj} \right\|^\alpha \pi_j(\theta) \exp\left\{ \varphi_{Tj}(\theta) - \varphi_{Tj}(\theta_j^0) \right\} \\
& \times \exp\left\{ -\frac{1}{2\varkappa_{jT}} \nabla\varphi_{Tj}(\theta_j^0)' J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0) \right\} d\theta
\end{aligned}$$

which in turn, is bounded by

$$\varkappa_{jT}^{(\alpha+1)/2} C K_n \int_{\left\|\theta-\xi_{Tj}\right\|>\delta/2} (1 + \|\theta\|^\alpha) \pi_j(\theta) \exp\left(\varphi_{Tj}(\theta) - \varphi_{Tj}(\theta_j^0) \right) d\theta \tag{66}$$

with $K_n := \exp(-\frac{1}{2\varkappa_{jT}} \nabla\varphi_{Tj}(\theta_j^0)' J_{Tj}(\theta_j^0)^{-1} \nabla\varphi_{Tj}(\theta_j^0)) = O_p(1)$, because $\xi_{Tj} = O_p(1)$. By Assumption 2, $\forall\delta>0$, $\exists\varepsilon>0$ such that

$$\lim_{n\to\infty} P\left\{\sup_{\left\|\theta_j-\theta_j^0\right\|>\delta/2} \exp\left(\varphi_{Tj}(\theta_j)-\varphi_{Tj}(\theta_j^0)\right)\leq \exp(-\varepsilon\varkappa_{jT})\right\}=1$$

so that (66) is bounded by

$$O_p(1)\varkappa_{jT}^{(\alpha+1)/2}\exp\left(-\varepsilon\varkappa_{jT}\right)\int_{\Theta}\left\|\theta\right\|^{\alpha}\pi_j(\theta)d\theta=o_p(1)$$

by boundedness of Θ and continuity of $\pi_j(.)$ in Assumption 1 and Assumption 4 respectively.

b) follows directly since the Gaussian density has infinitely many moments and $\varkappa_{jT}\to\infty$.

(iii) Area $\mathbb{H}_{Tj}^{(3)}$. By integrability of Gaussian density functions, we can choose M to be large enough to make the term in A_{1n} $\exp\left\{-\frac{1}{2}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}\right\}$ arbitrarily small. So, it is sufficient to show that for all $\varepsilon>0$, there exists M such that the remaining term

$$\liminf_{T\to\infty} P\left\{\int_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\|h\|^{a}\left|\pi_j(h/\sqrt{\varkappa_{jT}}+\xi_T)\exp(\omega_{Tj}(h))\right|dh<\varepsilon\right\}\geq 1-\varepsilon. \tag{67}$$

By (56),

$$e^{\omega(h_{Tj})}\leq e^{-\frac{1}{2}h'_{Tj}J_{Tj}(\theta_j^0)h_{Tj}+\left|R_{Tj}\left(\xi_T+h_{Tj}/\sqrt{\varkappa_{jT}}\right)\right|}. \tag{68}$$

Moreover, δ in the partition of $\mathbb{H}_{Tj}$ is chosen to satisfy (48): by (64) and (48), for all $\varepsilon>0$ and $\eta>0$, there exists a $\delta>0$ and $M>0$, such that

$$\begin{aligned}
1-\varepsilon &\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\frac{\left|R_{Tj}(h/\sqrt{\varkappa_{jT}}+\xi_{Tj})\right|}{\left\|h+\sqrt{\varkappa_{jT}}(\xi_{Tj}-\theta_j^0)\right\|^2}\leq\frac{1}{4}\eta\right\}\\
&\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\frac{\left|R_{Tj}(h/\sqrt{\varkappa_{jT}}+\xi_{Tj})\right|}{2\left\|h\right\|^2+2\left\|\sqrt{\varkappa_{jT}}(\xi_{Tj}-\theta_j^0)\right\|^2}\leq\frac{1}{4}\eta\right\}\\
&\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\frac{\left|R_{Tj}(h/\sqrt{\varkappa_{jT}}+\xi_{Tj})\right|}{\left\|h\right\|^2+C^2}\leq\frac{1}{2}\eta\right\}\\
&\leq \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\frac{\left|R_{Tj}(h/\sqrt{\varkappa_{jT}}+\xi_{Tj})\right|}{\left\|h\right\|^2\left(1+\frac{C^2}{M^2}\right)}\leq\frac{1}{2}\eta\right\}\\
&= \liminf_{T\to\infty} P\left\{\sup_{M<\|h\|\leq\delta\sqrt{\varkappa_{jT}}}\frac{\left|R_{Tj}(h/\sqrt{\varkappa_{jT}}+\xi_{Tj})\right|}{\left\|h\right\|^2}\leq\frac{1}{2}B\eta\right\}
\end{aligned} \tag{69}$$

for all $\varepsilon > 0$ and all $\eta > 0$ and some $C > 0$, with $B = 1 + C^2/M^2$. The equality in (69) follows since

$$\sqrt{\varkappa_{jT}}(\xi_{Tj} - \theta_j^0) = \frac{1}{\sqrt{\varkappa_{jT}}} J_{Tj}(\theta_j^0)^{-1} \nabla \varphi_{Tj}(\theta_j^0) = O_p(1)$$

so $\liminf_{T\to\infty} P\left\{\left\|\sqrt{\varkappa_{jT}}(\xi_{Tj} - \theta_j^0)\right\| \leq C\right\} = 1$ for some $C > 0$. Choosing $\eta = (2B)^{-1} \lambda_*$ (with λ_* defined in (47)) in (69) yields

$$\liminf_{T\to\infty} P\left\{\left|R_{Tj}(h_{Tj}/\sqrt{\varkappa_{jT}} + \xi_{Tj})\right| \leq \frac{1}{4} \|h_{Tj}\|^2 \lambda_* \right\} > 1 - \varepsilon. \tag{70}$$

Going back to (68), we want to show that

$$\liminf_{T\to\infty} P\left(e^{\omega(h_{Tj})} \leq C e^{-\frac{1}{4} h'_{Tj} J_{Tj}(\theta_j^0) h_{Tj}}\right) \geq 1 - \varepsilon.$$

Note that, by the identity $\min_{\|x\|=1} x'Ax/\|x\|^2 = \lambda_{\min}(A)$ for any symmetric A, we obtain

$$\begin{aligned}
-\frac{1}{4} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj} &\geq -\frac{1}{2} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj} + \frac{1}{4} \lambda_{\min}\left(J_{Tj}\left(\theta_j^0\right)\right) \|h_{Tj}\|^2 \\
&\geq -\frac{1}{2} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj} + \frac{1}{4} \lambda_* \|h_{Tj}\|^2
\end{aligned}$$

for all but finitely many T. Denoting

$$\begin{aligned}
G_{1T} &= -\frac{1}{2} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj} + \left|R_{Tj}\left(\xi_T + h_{Tj}/\sqrt{\varkappa_{jT}}\right)\right| \\
G_{2T} &= \frac{1}{4} \lambda_* \|h_{Tj}\|^2 - \left|R_{Tj}\left(\xi_T + h_{Tj}/\sqrt{\varkappa_{jT}}\right)\right|,
\end{aligned}$$

the above inequality implies that

$$\begin{aligned}
\liminf_{T\to\infty} P\left(e^{\omega(h_{Tj})} \leq C e^{-\frac{1}{4} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj}}\right) &\geq \liminf_{T\to\infty} P\left(e^{\omega(h_{Tj})} \leq C e^{-\frac{1}{2} h'_{Tj} J_{Tj}\left(\theta_j^0\right) h_{Tj} + \frac{1}{4}\lambda_* \|h_{Tj}\|^2}\right) \\
&= \liminf_{T\to\infty} P\left(e^{\omega(h_{Tj})} \leq C e^{G_{1T}} e^{G_{2T}}\right) \\
&\geq \liminf_{T\to\infty} P\left(e^{\omega(h_{Tj})} \leq C e^{G_{1T}} e^{G_{2T}} \middle| e^{G_{2T}} \geq 1\right) P\left(e^{G_{2T}} \geq 1\right) \\
&= \liminf_{T\to\infty} P\left(e^{G_{2T}} \geq 1\right) \\
&\geq 1 - \varepsilon
\end{aligned}$$

by (70), where we have used the fact that

$$P\left(e^{\omega(h_{Tj})} \leq Ce^{G_{1T}}e^{G_{2T}}\middle| e^{G_{2T}} \geq 1\right) = 1$$

for any $C \geq 1$ by (68) and (70). This completes the proof of Proposition 1.

7.1.2 Proof of Proposition 2

We have a linear regression model with parameters varying over time:

$$y_t = x_t\beta_t + \varepsilon_t.$$

where x_t is a $1 \times (k+1)$ vector of exogenous fixed regressors and β_t is a $(k+1) \times 1$ vector of time varying parameters, possibly including a time varying intercept term. ε_t are independent normally distributed mean zero disturbances with a variance σ_t^2, also indexed by time.

The weighted likelihood of the sample $Y := (y_1, ..., y_T)'$ at each point in time j is given by

$$L_j(Y|\beta_j, \lambda_j, X) = (2\pi)^{-S_j/2}\lambda_j^{Sj/2}\exp\left\{-\frac{\lambda_j}{2}\sum_{t=1}^{T}\vartheta_{jt}(y_t - x_t\beta_j)^2\right\}$$

where $S_j = \sum_{t=1}^{T}\vartheta_{jt} = \varkappa_{jT}$ and $\varkappa_{jT}$ is defined in (6), or equivalently in a more compact form as

$$L_j(Y|\beta_j, \lambda_j, X) = (2\pi)^{-S_j/2}\lambda_j^{Wj/2}\exp\left\{-\frac{\lambda_j}{2}(Y - X\beta_j)'D_j(Y - X\beta_j)\right\} \tag{71}$$

where $Y := (y_1, ..., y_T)'$ is a $T \times 1$ vector, $X = (x_1', ..., x_T')'$ is a $T \times k$ matrix and D_j is a $T \times T$ diagonal weighting matrix, containing the kernel weights in its main diagnal, $D_j = diag(\vartheta_{j1}, ..., \vartheta_{jT})$.

Next, assume that β_j and λ_j have Normal-Gamma prior distribution for $j \in \{1, ..., T\}$:

$$\beta_j|\lambda_j \sim \mathcal{N}\left(\beta_{0j}, (\lambda_j\kappa_{0j})^{-1}\right), \quad \lambda_j \sim Ga(\alpha_{0j}, \gamma_{0j})$$

implying that

$$\begin{aligned} p(\beta_j, \lambda_j) \quad &= \quad \frac{\gamma_{0j}^{a_{j0}}}{\Gamma(\alpha_{0j})}(2\pi)^{-(k+1)/2}|\kappa_{0j}|^{\frac{1}{2}}\lambda_j^{\alpha_{0j}-1}\lambda_j^{(k+1)/2} \\ &\quad \times\exp\left\{-\frac{\lambda_j}{2}((\beta_j - \beta_{0j})'\kappa_{0j}(\beta_j - \beta_{0j}) + 2\gamma_{0j})\right\} \end{aligned} \tag{72}$$

Combining the local likelihood function of the sample in (71) with the prior in (72) yields the following joint local quasi-posterior density, $p(\beta_j, \lambda_j|Y, X) \propto$

$$\lambda_j^{\alpha_{0j}-1+Sj/2+(k+1)/2} \exp\left\{-\frac{\lambda_j}{2}\left[2\gamma_{0j} + (\beta_j - \beta_{0j})'\kappa_{0j}(\beta_j - \beta_{0j}) + (Y - X\beta_j)'D_j(Y - X\beta_j)\right]\right\}.$$

It is useful to employ in the expression in the exponent above the following identity:

$$u'Au - 2a'u = (u - A^{-1}a)'A(u - A^{-1}a) - a'A^{-1}a \tag{73}$$

for any vectors u and a and a symmetric p.d. matrix A. In particular,

$$\begin{aligned} & -\frac{\lambda_j}{2}\left[2\gamma_{0j} + \beta_j{}'(X'D_jX + \kappa_{0j})\beta_j - 2(\beta_{0j}{}'\kappa_{0j} + Y'D_jX)\beta_j + \beta_{0j}{}'\kappa_{0j}\beta_{0j} + Y'D_jY\right] \\ = & -\frac{\lambda_j}{2}\left[2\widetilde{\gamma}_j + (\beta_j - \widetilde{\beta}_j)'\widetilde{\kappa}_j(\beta_j - \widetilde{\beta}_j)\right] \end{aligned}$$

and hence

$$p(\beta_j, \lambda_j|Y, X) \propto \lambda_j^{\alpha_{0j}+Sj/2-1+(k+1)/2} \exp\left(-\frac{\lambda_j}{2}\left[2\widetilde{\gamma}_j + (\beta_j - \widetilde{\beta}_j)'\widetilde{\kappa}_j(\beta_j - \widetilde{\beta}_j)\right]\right)$$

which is in fact of Normal-Gamma form:

$$\begin{aligned} \beta_j|X, Y, D_j, \lambda_j &\sim N\left(\widetilde{\beta}_j, (\lambda_j\widetilde{\kappa}_j)^{-1}\right) \\ \lambda_j|X, Y, D_j &\sim Ga(\widetilde{\alpha}_j, \widetilde{\gamma}_j) \end{aligned}$$

with parameters:

$$\begin{gathered} \widetilde{\beta}_j = \widetilde{\kappa}_j^{-1}(X'D_jX\hat{\beta}_j + \kappa_{0j}\beta_{0j}) \\ \widetilde{\kappa}_j = \kappa_{0j} + X'D_jX, \quad \tilde{\alpha}_j = \alpha_{0j} + S_j/2 \\ \widetilde{\gamma}_j = \gamma_{0j} + \frac{1}{2}\left(Y'D_jY - \widetilde{\beta}_j'\widetilde{\kappa}_j\widetilde{\beta}_j + \beta_{0j}'\kappa_{0j}\beta_{0j}\right) \end{gathered}$$

which completes the proof of Proposition 2.

7.1.3 Proof of Proposition 3

To obtain the marginal quasi-posterior of β_j, integrate out the joint quasi posterior given in () with respect to λ :

$$
\begin{aligned}
p(\beta_j|Y,X,D_j) &= \textstyle\int_0^\infty p(\beta_j,\lambda|Y,X)d\lambda \\
&= C_{4j}^{-1}\textstyle\int_0^\infty \lambda^{\tilde{\alpha}_j+\frac{k+1}{2}} \exp\left\{-\lambda\left[\tilde{\gamma}_j-\frac{1}{2}(\beta_j-\tilde{\beta}_j)'\tilde{\kappa}_j(\beta_j-\tilde{\beta}_j)\right]\right\}d\lambda
\end{aligned}
$$

After change of variables, $x=\lambda(\tilde{\gamma}_j-\dfrac{1}{2}(\beta_j-\tilde{\beta}_j)'\tilde{\kappa}_j(\beta_j-\tilde{\beta}_j))$, we obtain:

$$
\begin{aligned}
&= C_{4j}^{-1}\textstyle\int_0^\infty x^{(\tilde{\alpha}_j+\frac{k+1}{2})-1}\exp\{-x\}\left(\tilde{\gamma}_j-\frac{1}{2}(\beta_j-\tilde{\beta}_j)'\tilde{\kappa}_j(\beta_j-\tilde{\beta}_j)\right)^{-\tilde{\alpha}_j-\frac{k+1}{2}}dx \\
p(\beta_j|Y,X) &= \frac{\tilde{\gamma}_j^{\tilde{\alpha}_j}\,|\tilde{\kappa}_j|^{\frac{1}{2}}}{\Gamma(\tilde{\alpha}_j)(2\pi)^{\frac{k+1}{2}}}\left(\tilde{\gamma}_j-\frac{1}{2}(\beta_j-\tilde{\beta}_j)'\tilde{\kappa}_j(\beta_j-\tilde{\beta}_j)\right)^{-\tilde{\alpha}_j-\frac{k+1}{2}}\Gamma(\tilde{\alpha}_j+\frac{k+1}{2}) \\
&= \frac{\Gamma(\frac{2\tilde{\alpha}_j+k+1}{2})}{(2\tilde{\alpha}_j\pi)^{(k+1)/2}\tilde{\alpha}_j^{-(k+1)/2}\tilde{\gamma}_j^{(k+1)/2}\,|\tilde{\kappa}_j|^{-1/2}\,\Gamma(\tilde{\alpha}_j)} \\
&\times\left(1-\frac{2\tilde{\alpha}_j}{2\tilde{\gamma}_j 2\tilde{\alpha}_j}(\beta_j-\tilde{\beta}_j)'\tilde{\kappa}_j(\beta_j-\tilde{\beta}_j)\right)^{-(\tilde{\alpha}_j+\frac{k+1}{2})}
\end{aligned}
$$

Hence $p(\beta_j|Y)$ is the density of a Student's t-distribution of the form

$$
T_v(\mu,\sigma^2)=\frac{\Gamma(\frac{v+p}{2})}{\Gamma(\frac{v}{2})v^{\frac{p}{2}}\pi^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}\left(1-\frac{1}{v\sigma^2}(\beta-\mu)^2\right)^{\frac{v+p}{2}}}
$$

where $v=2\tilde{\alpha}_j$, $\mu=\tilde{\beta}_j$ and $\Sigma=\frac{\tilde{\gamma}_j}{\tilde{\alpha}_j}\tilde{\kappa}_j^{-1}$. Therefore, we have that $\beta_j|Y,X\sim T_{2\alpha_M}\left(\tilde{\beta}_j,\frac{\tilde{\gamma}_j}{\tilde{\alpha}_j}\tilde{\kappa}_j^{-1}\right)$ as required.

7.1.4 Proof of Proposition 4

We have that the $M\times 1$ dimensional vector y_t is generated by a time varying parameter VAR model of lag order k:

$$
y_t = B_{0t}+\sum\nolimits_{p=1}^{k}B_{pt}y_{t-p}+\varepsilon_t, \tag{74}
$$

where B_{0t} is an $M\times 1$ vector of time varying intercepts, and B_{pt} is an $M\times M$ matrix of time varying autoregressive coefficients for lag $p=1,...,k$. The error term, ε_t, is an $M\times 1$ vector of normally distributed zero mean random variables, with a positive definite symmetric $M\times M$

contemporaneous drifting covariance matrix R_t^{-1}, so that $\varepsilon_t = R_t^{-1/2}\eta_t$ where $\eta_t \sim NID(0_M, I_M)$. In addition, let $x_t := (1, y'_{t-1}, ..., y'_{t-k})$ be a $1 \times (Mk+1)$ vector and $B_t := (B_{0t}, B_{1t}, ..., B_{kt})$ be an $M \times (Mk+1)$ matrix. Then, (74) can be written as $y_t = B_t x'_t + \varepsilon_t$ and after vectorising, $y_t = (I_M \otimes x_t)\beta_t + R_t^{-1/2}\eta_t$, where $\beta_t := vec(B'_t)$ is an $M(Mk+1) \times 1$ vector.

The weighted likelihood of the sample $(y_1, ..., y_T)$ at each point in time j is given by

$$L_j(y|\beta_j, R_j, X) = (2\pi)^{-MS_j/2} |R_j|^{S_j/2} e^{-\frac{1}{2}\sum_{t=1}^{T} \vartheta_{jt}(y_t - (I_M \otimes x_t)\beta_j)' R_j (y_t - (I_M \otimes x_t)\beta_j)} \tag{75}$$

where $S_j = \sum_{t=1}^{T} \vartheta_{jt}$ and the kernel weights ϑ_{jt} are defined in (9). Denote by $Y = (y_1, ..., y_T)'$ a $T \times M$ matrix of stacked vectors $y'_1, ..., y'_T$ and define $y = vec(Y)$ as a $TM \times 1$ vector. Define $E = (\varepsilon_1, ..., \varepsilon_T)'$ implying that $\varepsilon := vec(E)$ is a $TM \times 1$ vector. Let X be a $T \times Mk + 1$ matrix defined as $X = (x'_1, ..., x'_T)'$. Then the weighted likelihood can be written in a more compact form as:

$$L_j(y|\beta_j, R_j, X) \propto |R_j|^{S_j/2} \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta_j)'(R_j \otimes D_j)(y - (I_M \otimes X)\beta_j)\right\}$$

where $D_j := diag(\vartheta_{j1}, ..., \vartheta_{jT})$ for $j \in \{1, ..., T\}$.

Next, specify a prior for β_j and R_j that has Normal-Wishart distribution:

$$\begin{aligned} p(\beta_j, R_j) \quad &\propto \quad |R_j|^{(Mk+1)/2} \exp\left\{-\frac{1}{2}(\beta_j - \beta_{0j})'(R_j \otimes \kappa_{0j})(\beta_j - \beta_{0j})\right\} \\ &\quad \times |R_j|^{\frac{\alpha_{0j} - M - 1}{2}} \exp\left\{-\frac{1}{2}tr(\gamma_{0j} R_j)\right\} \end{aligned}$$

where β_{0j} is a $(MK+1)M \times 1$ vector of prior means, κ_{0j} is a $(Mk+1) \times (Mk+1)$ positive definite symmetric precision matrix, α_{0j} is a scalar scale parameter of the Wishart distribution, γ_{0j} is a positive definite symmetric matrix, $tr(.)$ denotes the trace operator and $\propto$ denotes proportionality upto a constant. Combining the weighted likelihood with the prior, we obtain the local quasi-posterior density,

$$\begin{aligned} p(\beta_j, R_j|Y, X) \quad &\propto \quad \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta_j)'(R_j \otimes D_j)(y - (I_M \otimes X)\beta_j)\right\} |R_j|^{S_j/2} |R_j|^{(Mk+1)/2} \\ &\quad \times |R_j|^{\frac{\alpha_{0j} - M - 1}{2}} \exp\left\{-\frac{1}{2}(\beta_j - \beta_{0j})'(R_j \otimes \kappa_{0j})(\beta_j - \beta_{0j}) - \frac{1}{2}tr(\gamma_{0j} R_j)\right\}. \end{aligned} \tag{76}$$

A direct application of identity (73) in the expression in the exponential term in (76) yields:

$$
\begin{aligned}
&(y-(I_M\otimes X)\beta_j)'(R_j\otimes D_j)(y-(I_M\otimes X)\beta_j)+(\beta_j-\beta_{0j})'(R_j\otimes\kappa_{0j})(\beta_j-\beta_{0j})\\
&=y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}+\beta_j'(R_j\otimes(X'D_jX+\kappa_{0j}))\beta_j-2[R_j\otimes\left(XD_jy+\kappa_{0j}\beta_{0j}\right)]'\beta_j\\
&=y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}+\left(\beta_j-\widetilde{\beta}_j\right)'(R_j\otimes\widetilde{\kappa}_j)\left(\beta_j-\widetilde{\beta}_j\right)-\widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j \qquad (77)
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{\beta}_j &= \left(I_M\otimes\widetilde{\kappa}_j^{-1}\right)\left[(I_M\otimes X'D_jX)\hat{\beta}_j+(I_M\otimes\kappa_{0j})\beta_{0j}\right], \qquad (78)\\
\widetilde{\kappa}_j &= \kappa_{0j}+X'D_jX.
\end{aligned}
$$

It remains to study the remaining terms in (77):

$$
\exp\left\{-\frac{1}{2}\left[y'(R_j\otimes D_j)y+\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j}-\widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j+tr(\gamma_{0j}R_j)\right]\right\}.
$$

The first three terms will receive the same treatment. For example, the first term can be written as:

$$
\begin{aligned}
y'(R_j\otimes D_j)y &= \left[(I_M\otimes D_j^{1/2})vec(Y)\right]'(R_j\otimes I_T)\left[(I_M\otimes D_j^{1/2})vec(Y)\right]\\
&= vec(D_j^{1/2}Y)'(R_j\otimes I_T)vec(D_j^{1/2}Y)\\
&= tr(D_j^{1/2}YR_jY'D_j^{1/2})=tr(Y'D_jYR_j) \qquad (79)
\end{aligned}
$$

where the second line is obtained using the equality $vec(ABC)=(C'\otimes A)vec(B)$ and the third line uses the equality $tr(ABC)=vec(A')'(I\otimes B)vec(C)=vec(A)'(B\otimes I)vec(C')$. Similarly,

$$
\begin{aligned}
\beta_{0j}'(R_j\otimes\kappa_{0j})\beta_{0j} &= tr(B_{0j}\kappa_{0j}B_{0j}'R_j), \quad \text{and} \qquad (80)\\
\widetilde{\beta}_j'(R_j\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j &= tr(\widetilde{B}_j\widetilde{\kappa}_j\widetilde{B}_j'R_j), \qquad (81)
\end{aligned}
$$

where $\beta_{0j}=vec(B_{0j}')$ and $\widetilde{\beta}_j=vec(\widetilde{B}_j')$. After combining (77), (79),(80) and (81) the quasi-posterior in (76) can be written as: $p(\beta_j,R_j|Y,X)\propto$

$$
|R_j|^{(Mk+1)/2}|R_j|^{\frac{\widetilde{\alpha}_j-M-1}{2}}\exp\left\{-\frac{1}{2}\left[(\beta_j-\widetilde{\beta}_j)'(R_j\otimes\widetilde{\kappa}_j)(\beta_j-\widetilde{\beta}_j)+tr\left(\widetilde{\gamma}_jR_j\right)\right]\right\} \qquad (82)
$$

which is a Normal-Wishart density with parameters $\widetilde{\beta}_j, \widetilde{\kappa}_j, \widetilde{\alpha}_j$ and $\widetilde{\gamma}_j$, where $\widetilde{\beta}_j$ and $\widetilde{\kappa}_j$ are defined above in (78) and

$$\widetilde{\alpha}_j = \alpha_{0j} + S_j, \quad \widetilde{\gamma}_j = \gamma_{0j} + Y' D_j Y + B_{0j}\kappa_{0j}B'_{0j} - \widetilde{B}_j \widetilde{\kappa}_j \widetilde{B}'_j,$$

which proves Proposition 4.

7.1.5 Proof of Proposition 5

To obtain the marginal quasi-density of the parameter vector β_j, we integrate the joint quasi-posterior distribution, $p(\beta_j, R|Y, X)dR$ in (82) over the $M(M+1)/2$ distinct variables in R_j,

$$\begin{aligned} p(\beta_j|Y,X) &= \int_{R>0} p(\beta_j, R|Y,X)dR \\ &= C_j \int_{R>0} \exp\left\{-\frac{1}{2}(\beta_j - \widetilde{\beta}_j)'(R \otimes \widetilde{\kappa}_j)(\beta_j - \widetilde{\beta}_j)\right\} \\ &\quad \times |R|^{(Mk+1)/2} |R|^{\frac{\widetilde{\alpha}_j - M - 1}{2}} \exp\left\{-\frac{1}{2} tr\left(\widetilde{\gamma}_j R\right)\right\} dR \end{aligned} \tag{83}$$

where

$$C_j = \frac{(2\pi)^{-M(Mk+1)/2} |\widetilde{\kappa}_j|^{M/2}}{2^{\frac{1}{2}M\widetilde{\alpha}_j} \left|\widetilde{\gamma}_j\right|^{-\widetilde{\alpha}_j/2} \Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}.$$

Note that the first exponential term in (83) can be written as

$$\begin{aligned} & \exp\left\{-\frac{1}{2}(\beta_j - \widetilde{\beta}_j)'(R \otimes \widetilde{\kappa}_j)(\beta_j - \widetilde{\beta}_j)\right\} \\ = & \exp\left\{-\frac{1}{2} vec\left[\widetilde{\kappa}_j^{1/2}(B_j - \widetilde{B}_j)'\right]' (R \otimes I_{Mk+1}) vec\left[\widetilde{\kappa}_j^{1/2}(B_j - \widetilde{B}_j)'\right]\right\} \\ = & \; etr\left\{-\frac{1}{2}(B_j - \widetilde{B}_j)\widetilde{\kappa}_j(B_j - \widetilde{B}_j)'R\right\}, \end{aligned}$$

where $etr(\cdot)$ is the exponential trace operator, so that

$$p(\beta_j|Y,X) = C_{5j} \int_{R>0} |R|^{\frac{\widetilde{\alpha}_j - M - 1 + Mk + 1}{2}} etr\left\{-\frac{1}{2}\left[(B_j - \widetilde{B}_j)\widetilde{\kappa}_j(B_j - \widetilde{B}_j)' + \widetilde{\gamma}_j\right] R\right\} dR. \tag{84}$$

By Theorem 2.1.11 in Muirhead (2005), for a $n \times n$ symmetric positive definite matrix R, with $\mathrm{Re}(\Psi) > 0$ and $\mathrm{Re}(\delta) > 0$, the following holds:

$$\int_{R>0} |R|^{\delta - \frac{n-1}{2}} \exp\left\{-\frac{1}{2} tr(\Psi^{-1}R)\right\} d(R) = \Gamma_n(\delta)|\Psi|^{\delta} 2^{n\delta}.$$

Applying the theorem in (84), we have that

$$\begin{aligned}
p(\beta_j|Y,X) &= C_{5j}\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)\left|(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+\widetilde{\gamma}_j\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}}2^{M\frac{\widetilde{\alpha}_j+Mk+1}{2}} \\
&= \frac{(2\pi)^{-M(Mk+1)/2}\left|\widetilde{\kappa}_j\right|^{M/2}}{2^{-\frac{1}{2}M\widetilde{\alpha}_j}\left|\widetilde{\gamma}_j\right|^{-\widetilde{\alpha}_j/2}\Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)\left|\widetilde{\gamma}_j\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}} \\
&\quad\times\left|\widetilde{\gamma}_j^{-1}(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+I_M\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}}2^{M\frac{\widetilde{\alpha}_j+Mk+1}{2}}
\end{aligned}$$

$$= \pi^{-M(Mk+1)/2}\left|\widetilde{\kappa}_j\right|^{M/2}\left|\widetilde{\gamma}_j\right|^{-\frac{Mk+1}{2}}\frac{\Gamma_M\left(\frac{\widetilde{\alpha}_j+Mk+1}{2}\right)}{\Gamma_M\left(\frac{\widetilde{\alpha}_j}{2}\right)}\left|\widetilde{\gamma}_j^{-1}(B_j-\widetilde{B}_j)\widetilde{\kappa}_j(B_j-\widetilde{B}_j)'+I_M\right|^{-\frac{\widetilde{\alpha}_j+Mk+1}{2}}.$$

Recall the definition of a matrix variate t-distribution density of a $p\times m$ matrix B:

$$\begin{aligned}
p(B;\nu,M,\Sigma,\Omega) &= \frac{\Gamma_p(\frac{1}{2}(\nu+m+p-1))}{\pi^{\frac{1}{2}mp}\Gamma_p(\frac{1}{2}(\nu+p-1))}\left|\Sigma\right|^{-\frac{1}{2}m}\left|\Omega\right|^{-\frac{1}{2}p} \\
&\quad\times\left|I_p+\Sigma^{-1}(B-M)\Omega^{-1}(B-M)'\right|^{-\frac{1}{2}(\nu+m+p-1)},
\end{aligned}$$

and note that $p(\beta_j|Y,X)$ is a vectorised counterpart of the above definition with parameters:

$$m=Mk+1,\quad p=M,\quad \nu=\widetilde{\alpha}_j-Mk,\quad M=\widetilde{B}_j,\quad \Omega^{-1}=\widetilde{\kappa}_j,\quad \Sigma=\widetilde{\gamma}_j.$$

We therefore have

$$\beta_j|Y,X=vec(B_j')|Y,X\sim T_{\widetilde{\alpha}_j-Mk}\left(vec(\widetilde{B}_j'),\frac{\widetilde{\gamma}_j\otimes\widetilde{\kappa}_j^{-1}}{\widetilde{\alpha}_j-Mk-2}\right),$$

which proves Proposition 5.

7.1.6 Proof of Proposition 6

Assume that the variance covariance matrix $R_t^{-1/2}$ in equation (13) is known. Then, transform the model in the following way:

$$R_t^{1/2}y_t=R_t^{1/2}(I_M\otimes x_t)\beta_t+\eta_t,\quad \eta_t\sim NID(0_M,I_M).$$

Next, specify a prior for β_t of the form:

$$\beta_j \sim \mathcal{N}\left(\beta_{0j}, V_{0j}^{-1}\right) \text{ for } \; j \in \{1, ..., T\}.$$

Combining the prior with the local likelihood in (75) yields a quasi-posterior density for β_j of the form

$$p(\beta_j|Y, X, R_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[\sum\nolimits_{t=1}^{T} \vartheta_{jt}(\widetilde{y}_t - \widetilde{x}_t\beta_j)'(\widetilde{y}_t - \widetilde{x}_t\beta_j) + (\beta_j - \beta_{0j})'V_{0j}(\beta_j - \beta_{0j})\right]\right\}$$

where $\widetilde{x}_t = R_t^{1/2}(I_M \otimes x_t)$ and $\widetilde{y}_t = R_t^{1/2} y_t$. A direct application of identity (73) implies that the conditional quasi-posterior of β_j is $\mathcal{N}\left(\widetilde{\beta}_j, \widetilde{V}_j^{-1}\right)$:

$$p(\beta_j|Y, X, R_{1:T}) \propto \exp\left[-\frac{1}{2}\left(\beta_j - \widetilde{\beta}_j\right)' \widetilde{V}_j \left(\beta_j - \widetilde{\beta}_j\right)\right]$$

with posterior parameters

$$\widetilde{\beta}_j = \widetilde{V}_j^{-1}\left[\sum\nolimits_{t=1}^{T} \vartheta_{jt}\widetilde{x}_t'\widetilde{y}_t + V_{0j}\beta_{0j}\right], \quad \widetilde{V}_j = V_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt}\widetilde{x}_t'\widetilde{x}_t$$

which proves Proposition 6.

7.1.7 Proof of Proposition 7

Assume that β_t is known in the model in equation (13). Write the model as:

$$\varepsilon_t := y_t - (I_M \otimes x_t)\beta_t = R_t^{-1/2}\eta_t, \;\; \eta_t \sim NID(0_M, I_M),$$

and specify a Wishart $W(\alpha_{0j}, \gamma_{0j})$ prior density for R_t^{-1}. Combine the prior with the local likelihood (75) to obtain conditional quasi-posterior density for R_j of the form

$$p(R_j|Y, X, \beta_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[\sum\nolimits_{t=1}^{T} \vartheta_{jt}\varepsilon_t{}'R_j\varepsilon_t + tr(\gamma_{0j}R)\right]\right\} |R_j|^{\sum_{t=1}^{T}\vartheta_{jt}/2} |R_j|^{(Mk+\alpha_{0j}-M)/2}. \tag{85}$$

After stacking the observed ε_t as a $T \times M$ matrix $E := (\varepsilon_1, \varepsilon_2, ..., \varepsilon_T)'$, we can write $\sum_{t=1}^{T} \vartheta_{jt} \varepsilon_t{}' R_j \varepsilon_t = vec(E)'(R_j \otimes D_j) vec(E)$ where $D_j = diag(\vartheta_{jt})$. The calculation

$$
\begin{aligned}
& vec(E)'(R_j \otimes D_j)vec(E) \\
= \; & \left[(I_M \otimes D_j^{1/2})vec(E)\right]' (R_j \otimes I_T) \left[(I_M \otimes D_j^{1/2})vec(E)\right] \\
= \; & vec(D_j^{1/2}E)'(R_j \otimes I_T)vec(D_j^{1/2}E) \\
= \; & tr(D_j^{1/2} E R_j E' D_j^{1/2}) = tr(E' D_j E R_j)
\end{aligned}
$$

implies that the conditional quasi-posterior density for R_j in (85) is of $W(\widetilde{\alpha}_j, \widetilde{\gamma}_j)$ form:

$$
p(R_j | Y, X, \beta_{1:T}) \propto \exp\left\{-\frac{1}{2}\left[tr(\widetilde{\gamma}_j R)\right]\right\} |R_j|^{\frac{\widetilde{\alpha}_j - M - 1}{2}}
$$

with posterior parameters

$$
\widetilde{\alpha}_j = \alpha_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt}, \quad \widetilde{\gamma}_j = \gamma_{0j} + \sum\nolimits_{t=1}^{T} \vartheta_{jt} \varepsilon_t' \varepsilon_t.
$$

7.1.8 Proof of Proposition 8

To obtain the quasi-marginal likelihood of the sample Y, we integrate the local likelihood over all possible values of the parameters β_j and R_j :

$$
\begin{aligned}
p_j(Y) = & \int \int p(Y|\beta, R) p(\beta|R) p(R) d\beta dR \\
= (2\pi)^{-S_j/2} |\kappa_{0j}|^{(Mk+1)/2} & \frac{(2)^{-M\alpha_{0j}/2}}{\Gamma_M(\alpha_{0j}/2)} |\gamma_{0j}|^{\alpha_{0j}/2} \int |R|^{S_j/2} |R|^{\frac{\alpha_{0j}-M-1}{2}} etr\left(\frac{1}{2}\gamma_{0j}R\right) \\
\times \int (2\pi)^{-(Mk+1)/2} & \exp\left\{-\frac{1}{2}(y - (I_M \otimes X)\beta)'(R \otimes D_j)(y - (I_M \otimes X)\beta)\right\} |R|^{(Mk+1)/2} \\
& \times \exp\left\{-\frac{1}{2}(\beta - \beta_{0j})'(R \otimes \kappa_{0j})(\beta - \beta_{0j})\right\} d\beta dR
\end{aligned}
$$

Completing the square in the exponential term, by using idenity (73), we have

$$\begin{aligned} p_j(Y) &= (2\pi)^{-S_j/2} \left|\kappa_{0j}\right|^{(Mk+1)/2} \frac{(2)^{-M\alpha_{0j}/2}}{\Gamma_M(\alpha_{0j}/2)} \left|\gamma_{0j}\right|^{\alpha_{0j}/2} \int |R|^{S_j/2} |R|^{\frac{\alpha_{0j}-M-1}{2}} etr\left(\frac{1}{2}\gamma_{0j}R\right) \\ &\times \exp\left\{-\frac{1}{2}\left(y'(R\otimes D_j)y + \beta_{0j}'(R\otimes\kappa_{0j})\beta_{0j} - \widetilde{\beta}_j'(R\otimes\widetilde{\kappa}_j)\widetilde{\beta}_j\right)\right\} |\widetilde{\kappa}_j|^{-(Mk+1)/2} \\ &\underbrace{\times \int (2\pi)^{-(Mk+1)/2} |R\otimes\widetilde{\kappa}_j|^{(Mk+1)/2} \exp\left\{-\frac{1}{2}\left(\left(\beta-\widetilde{\beta}_j\right)'(R\otimes\widetilde{\kappa}_j)\left(\beta-\widetilde{\beta}_j\right)\right)\right\} d\beta dR}_{=1} \end{aligned}$$

where $\widetilde{\beta}_j$ and $\widetilde{\kappa}_j$ are defined in Proposition 4. After transforming remaining terms in the exponent above, using equations (79), (80) and (81), we have

$$\begin{aligned} p_j(Y) &= \frac{(2)^{(M-1)S_j/2} \left|\kappa_{0j}\right|^{(Mk+1)/2} \Gamma_M(\widetilde{\alpha}_j/2) \left|\gamma_{0j}\right|^{\alpha_{0j}/2}}{\pi^{S_j/2} |\widetilde{\kappa}_j|^{-(Mk+1)/2} \Gamma_M(\alpha_{0j}/2) \left|\widetilde{\gamma}_j\right|^{\widetilde{\alpha}_j/2}} \\ &\quad \underbrace{\times \int \frac{1}{\Gamma_M(\widetilde{\alpha}_j/2)} (2)^{-M\widetilde{\alpha}_j/2} \left|\widetilde{\gamma}_j\right|^{\widetilde{\alpha}_j/2} |R|^{\frac{\widetilde{\alpha}_j-M-1}{2}} etr\left\{-\frac{1}{2}\widetilde{\gamma}_j R\right\} dR}_{=1} \end{aligned}$$

where $\widetilde{\alpha}_j$ and $\widetilde{\gamma}_j$ are defined in Proposition 4, which completes the proof of Proposition 8.

7.1.9 Additional Algorithms

Homoscedastic BVAR model with mixture of time varying and time invariant slope parameters Consider a homoscedastic VAR model, which includes, in addition to the regressors with time varying coefficients, a subset of regressors with invariant parameters:

$$y_t = (I_M \otimes Z_t)\zeta + (I_M \otimes X_t)\beta_t + R^{-1/2}\eta_t, \quad \eta_t \sim \mathcal{N}(0, I_M)$$

where Z_t are fixed exogenous variables or lags of y_t.

Conditional on knowing R and ζ, redefine $\widehat{y}_t := R^{1/2}(y_t - (I_M \otimes Z_t)\zeta) = \widehat{x}_t\beta_t + \eta_t$ where $\widehat{x}_t = R^{1/2}(I_M \otimes X_t)$. Assuming a normal $\mathcal{N}(\beta_{0j}, V_{0j})$ prior for β_j, the conditional quasi-posterior of β_j is also normal by Proposition 6: $\beta_j|Y, X, Z, R^{-1}$, $\beta_j \sim \mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$, with parameters defined in (22). On the other hand, conditional on a draw from the posterior of β_j for each $j \in \{1, ..., T\}$, the model can be written as $\widetilde{y}_t = y_t - (I_M \otimes X_t)\beta_t = -(I_M \otimes Z_t)\zeta + R^{-1/2}\eta_t$. Hence, conditional on the time varying β_j, we have a standard linear Gaussian model with fixed parameters. Then, assuming a Normal-Wishart prior $\zeta, R \sim NW(\zeta_0, \lambda_0, a_0, \gamma_0)$ for R^{-1} and ζ, it follows that

$$\zeta, R|X, Y, Z, \beta_{1:T} \sim NW(\widetilde{\zeta}, \widetilde{\lambda}, \widetilde{a}, \widetilde{\gamma})$$

with parameters:

$$\begin{aligned} \widetilde{\zeta} &= I_M \otimes \left[(\lambda_0 + \sum_{t=1}^{T} Z_t' Z_t)^{-1} \left(\sum_{t=1}^{T} Z_t' \widetilde{y}_t + \lambda_0 \zeta_0 \right) \right] \\ \widetilde{\lambda} &= \lambda_0 + \sum_{t=1}^{T} Z_t' Z_t, \quad \widetilde{\alpha} = \alpha_0 + T, \quad \widetilde{\gamma} = \gamma_0 + \widetilde{y}_t' \widetilde{y}_t + S_0 \lambda_0 S_0' - \widetilde{S} \widetilde{\lambda} \widetilde{S}', \end{aligned} \tag{86}$$

where $\zeta_0 = vec(S_0')$ and $\widetilde{\zeta} = vec(\widetilde{S}')$.

Since the conditional posterior distribution of R^{-1} and ζ is of standard form and the conditional quasi-posterior of β_t have been established in Proposition 6, a Gibbs algorithm can be designed to approximate the joint posterior of R^{-1}, ζ and β_t , using the following steps.

Algorithm 3. **Step 1.** Initialise the algorithm with ζ^0 and $R^{-1,0}$.

For $i = 1, ..., N$ iterate between steps 2 and 3 below.

Step 2. For $j \in \{1, ..., T\}$ draw $\beta_j^i | Y, X, R^{-1,i-1}, \zeta^{i-1}$ from $\mathcal{N}(\widetilde{\beta}_j, \widetilde{V}_j)$ with posterior parameters in (22).

Step 3. Draw $R^{-1,i}, \zeta^i | Y, X, \beta_{1:T}^i$ from $NW(\widetilde{\zeta}, \widetilde{\lambda}, \widetilde{\alpha}, \widetilde{\gamma})$ with posterior parameters in (86).

Heteroscedastic BVAR model with mixture of time varying and time invariant slope parameter Next, we consider a heteroscedastic BVAR model which includes, in addition to heteroscedasticity and fixed slope coefficients ζ, a subset of regressors which enter the model with drifting parameters β_t

$$y_t = (I_M \otimes Z_t)\zeta + (I_M \otimes X_t)\beta_t + R_t^{-1/2} \eta_t, \quad \eta_t \sim \mathcal{N}(0, I_M) \tag{87}$$

where Z_t are fixed exogenous variables or lags of y_t. Conditional on ζ, the model can be written as

$$\widehat{y}_t = y_t - (I_M \otimes Z_t)\zeta = (I_M \otimes X_t)\beta_t + R_t^{-1/2} \eta_t.$$

Draws from the joint quasi-posterior of β_t and R_t can be obtained using the closed form expressions in Proposition 1, using $\widehat{y}_t$ instead of y_t; if a Normal-Wishart prior $NW(\beta_{0j}, \kappa_{0j}, \alpha_{0j}, \gamma_{0j})$

is assumed for β_j and R_j, then, the posterior is also Normal-Wishart $NW(\widetilde{\beta}_j, \widetilde{\kappa}_j, \widetilde{a}_j, \widetilde{\gamma}_j)$ with parameters given in (19).

On the other hand, conditional on the draw from the history of $\beta_{1:T}$ and $R_{1:T}$; we can define $y_t^* = R_t^{1/2}(y_t - (I_M \otimes X_t)\beta_t) = z_t^* \zeta + \eta_t$ with $z_t^* = R_t^{1/2}(I_M \otimes Z_t)$ and the model reduces to the standard linear Gaussian model with fixed parameters and known variance. Then, by assuming a normal prior $\mathcal{N}(\zeta_0, Q_0)$ for ζ, the conditional posterior is also Normal: $\zeta | Y, X, Z, R_{1:T}^{-1}, \beta_{1:T} \sim \mathcal{N}(\widetilde{\zeta}, \widetilde{Q})$ with parameters

$$\widetilde{\zeta} = \left(\sum_{t=1}^{T} z_t^{*\prime} z_t^* + Q_0^{-1} \right)^{-1} \left(\sum_{t=1}^{T} z_t^{*\prime} y_t^* + Q_0^{-1} \zeta_0 \right), \quad \widetilde{Q} = \left(\sum_{t=1}^{T} z_t^{*\prime} z_t^* + Q_0^{-1} \right)^{-1}. \tag{88}$$

Hence, the conditional posterior of ζ is of standard form and the conditional quasi-posterior of β_j and R_j^{-1} have been characterised in Proposition 1 and can be easily drawn from, the model (87) permits the use of a Gibbs algorithm with the following steps to approximate the joint posterior of $\beta_{1:T}, R_{1:T}$ and ζ.

Algorithm 4. **Step 1.** Initialise the algorithm with ζ^0.

For $i = 1, ..., N$ iterate between steps 2 and 3 below.

Step 2. For each $j \in \{1, ..., T\}$, draw $R_j^{-1,i}$ and β^i from $\beta_j, R_j | X, Y, Z, \zeta^{i-1} \sim NW(\widetilde{\beta}_j, \widetilde{\kappa}_j, \widetilde{a}_j, \widetilde{\gamma}_j)$, with posterior parameters defined in (19).

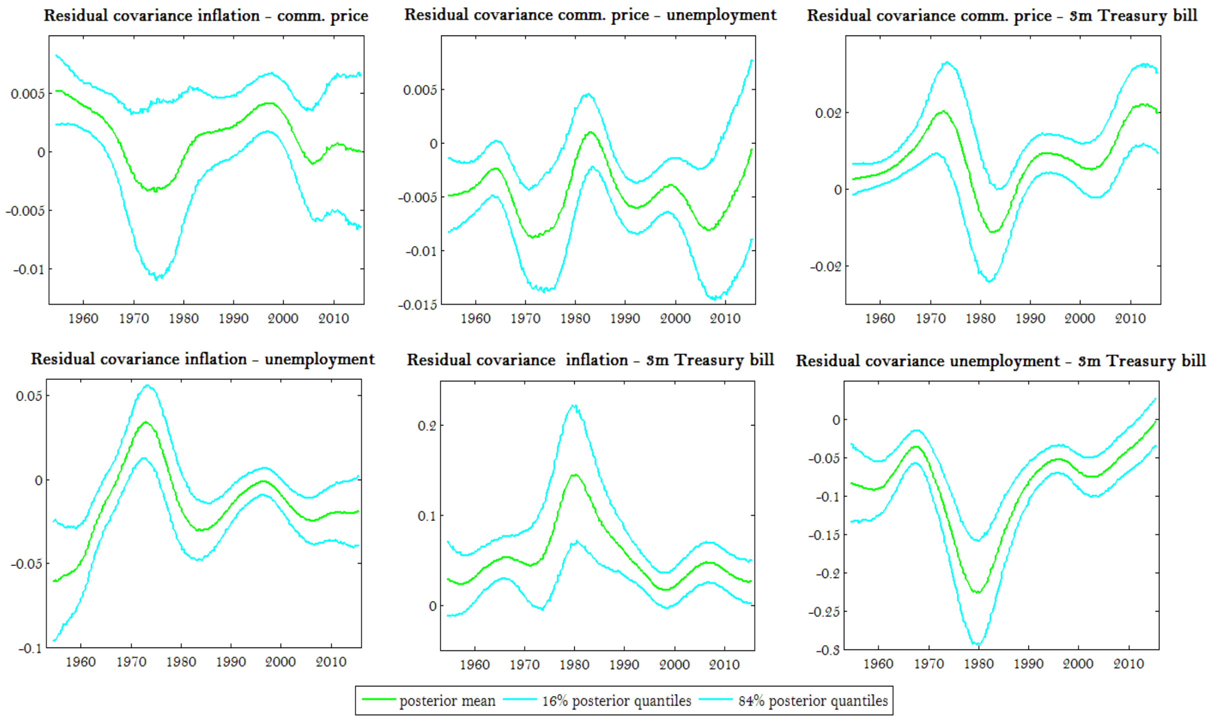
Step 3: Draw $\zeta^i | Y, X, Z, R_{1:T}^{-1,i}, \beta_{1:T}^i$ from $\mathcal{N}(\widetilde{\zeta}, \widetilde{Q})$ with posterior parameters defined in (88).

7.1.10 Additional results and data description

	Series Desription	Transformation
1	Real GDP	DLOG(GDP)*100
2	GDP Deflator	DLOG(GDPDEF)*100
3	Consumer Price Index	DLOG(CPI)*100
4	Business Investment	DLOG(INV)
5	Real Personal Consumption Expenditure	DLOG(CONS)*100
6	Civilian Unemployment Rate	D(UNRATE/4)
7	Industrial Production	DLOG(IP)*100
8	3-Month Treasury Bill: Secondary Market Rate	D(TBILL/4)
9	ISM Manufacturing: PMI Composite Index	DLOG(PMI)
10	ISM Manufacturing: New Orders Index	DLOG(NEWORDERS)
11	Average Hours	DLOG(AVERAGEHOURS)*100
12	Nonfarm Business Sector: Real Compensation Per Hour	DLOG(REALWAGE)*100
13	Producer Price Index	DLOG(PPI)*100
14	Personal Consumption Expenditures: Chain-type Price Index	DLOG(PCE)*100
15	Reuters/Jeffries-CRB Total Return Index (w/GFD extension)	DLOG(COMINDEX)
16	BAA Corporate Spread	D(CORPSPREAD)
17	NYSE Stock Market Capitalization	DLOG(STOCKCAP)
18	Industrial Production: Business Equipment	DLOG(IPBUSEQ)*100
19	Industrial Production: Consumer Goods	DLOG(IPCONGD)*100
20	Industrial Production: Durable Consumer Goods	DLOG(IPDCONGD)*100
21	Industrial Production: Durable Materials	DLOG(IPDMAT)*100
22	Industrial Production: Final Products (Market Group)	DLOG(IPFINAL)*100
23	Industrial Production: Final Products and Nonindustrial Supplies	DLOG(IPFPNSS)*100
24	Industrial Production: Manufacturing (SIC)	DLOG(IPMANSICS)*100
25	Industrial Production: Materials	DLOG(IPMAT)*100
26	Industrial Production: Nondurable Consumer Goods	DLOG(IPNCONGD)*100
27	Dow Jones Industrial Total returns index	DLOG(DOW)
28	ISM Manufacturing: Inventories Index	DLOG(INVENTORIES)
29	ISM Manufacturing: Supplier Deliveries Index	DLOG(SUPPLIERS)
30	ISM Manufacturing: PMI Composite Index	DLOG(NAPM)
31	ISM Manufacturing: PMI Employment index	DLOG(NAPMEI)
32	ISM Manufacturing: Production index	DLOG(NAPMPI)
33	ISM Manufacturing: Prices index	DLOG(NAPMPRI)
34	Civilian Employment	DLOG(EMPLOY)*100
35	All Employees: Construction	DLOG(USCONS)*100
36	All Employees: Financial Activities	DLOG(USFIRE)*100
37	All Employees: Good producing industries	DLOG(USGOOD)*100
38	All Employees: Government	DLOG(USGOVT)*100
39	All Employees: Trade and Transportation	DLOG(USTPU)*100
40	All Employees: retail trade	DLOG(USTRADE)*100
41	All Employees: wholesale trade	DLOG(USWTRADE)*100
42	All Employees: Durable Goods	DLOG(DMANEMP)*100
43	All Employees: Manufacturing	DLOG(MANEMP)*100
44	All Employees: Non-Durable Goods	DLOG(NDMANEMP)*100
45	All Employees: Service Providing industries	DLOG(SRVPRD)*100
46	Total Non-Farm Payrolls	DLOG(PAYEMS)*100
47	Real Personal Incomes excluding current transfers	DLOG(W875RX1)*100
48	Business Conditions Index	DLOG(BCI)
49	Real Imports	DLOG(IMPORTS)*100
50	Real Exports	DLOG(EXPORTS)*100
51	Real Government Spending	DLOG(REALGS)*100
52	Real Net Taxes	DLOG(REALTAX)
53	Number of Civilians unemployed for 15 weeks or over	DLOG(UEMP15OV)
54	Number of Civilians unemployed for 15 to 26 weeks	DLOG(UEMP15T26)
55	Number of Civilians unemployed 27 weeks and over	DLOG(UEMP27OV)
56	Number of Civilians unemployed for 5 to 14 weeks	DLOG(UEMP5TO14)
57	Number of Civilians unemployed for less than 5 weeks	DLOG(UEMPLT5)
58	Average Mean Duration of employment	DLOG(UEMPMEAN)
59	Average Weekly Hours of Production and Nonsupervisory Employees:Goods-Producing	DLOG(CES0600000007)*100
60	Average Hourly Earnings of Production and Nonsupervisory Employees:Goods-Producing	DLOG(CES0600000008)*100
61	Average Hourly Earnings of Production and Nonsupervisory Employees:Construction	DLOG(CES2000000008)*100
62	Average Weekly hours of Production and Nonsupervisory Employees:Manufacturing	DLOG(CES3000000008)*100
63	Average Hourly Earnings of Production and Nonsupervisory Employees:Manufacturing	DLOG(AWHMAN)*100
64	Civilian Labour Force	DLOG(CLF)*100
65	Civilian Participation rate	DLOG(CIVPART)*100
66	Nonfarm Business Sector: Unit Labor Cost	DLOG(ULC)*100
67	M2 Money Stock	DLOG(M2)*100
68	Total Consumer Credit Owned and Securitized, Outstanding	DLOG(CREDIT)*100
69	Commercial and Industrial Loans, All Commercial Banks	DLOG(BUSLOANS)*100
70	Real Estate Loans, All Commercial Banks	DLOG(REALLN)*100
71	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals	DLOG(PPICMM)*100
72	Producer Price Index: Crude Materials for Further Processing	DLOG(PPICRM)*100
73	Producer Price Index: Finished Consumer Goods	DLOG(PPIFCG)*100
74	Producer Price Index: Finished Goods	DLOG(PPIFGS)*100
75	Producer Price Index: Intermediate Materials: Supplies & Components	DLOG(PPIITM)*100
76	Consumer Price Index for All Urban Consumers: Apparel	DLOG(CPIAPPSL)*100
77	Consumer Price Index for All Urban Consumers: Medical Care	DLOG(CPIMEDSL)*100
78	Consumer Price Index for All Urban Consumers: All items less shelter	DLOG(CUUR0000SA0L2_D11)*100
79	10 year Govt Bond Yield minus 3 mth yield	D(GB10-TBILL)
80	6-month Treasury bill minus 3 mth yield	D(TBILL6-TBILL)
81	1 year Govt Bond Yield minus 3 mth yield	D(GB1-TBILL)
82	5 year Govt Bond Yield minus 3 mth yield	D(GB5-TBILL)
83	AAA Corporate Bond Spread	D(AAA-GB10)
84	S&P500 Total Return Index	DLOG(STOCK)
85	S&P500 P/E Ratio	D(PE)
86	US UK Exchange Rate	DLOG(DOLLARATE)
87	US Canada exchange rate	DLOG(CNRATE)*100

Large BVAR models in the forecasting exercise are estimated with all 87 variables, medium BVAR models include variables 1-17.

Figure 43: Off-diagonal covariance matrix elements over time



7.1.11 Additional Monte Carlo Results

This Section contains earlier Monte Carlo simulations. The DGP is given by

$$y_t = \beta_t + u_t, \quad u_t \sim NID(0, \sigma_t^2) \tag{89}$$

$$\beta_t = 1/\sqrt{t}\sum\nolimits_{i=1}^{t} \omega_t \tag{90}$$

$$\log \sigma_t^2 = 1/\sqrt{t}\sum\nolimits_{i=1}^{t} v_t \tag{91}$$

Equations (90) and (91) can be equivalently written as

$$\beta_t = \sqrt{(t-1)/t}\beta_{t-1} + \omega_t/\sqrt{t}$$

and

$$\log \sigma_t^2 = \sqrt{(t-1)/t}\log \sigma_{t-1}^2 + v_t/\sqrt{t}$$

respectively. The models that we compare are the QBLL estimator of with Normal-Gamma prior introduced in Section 2.3, a QBLL estimator with a uniform prior (which delivers an estimator equivalent to the frequentist estimator of Giraitis et al. (2014)), a state space model with volatility modelled as in Kim et al. (1998) and a particle filter. We present two sets of results: i) when the correct state equation is specified and ii) when instead we fit a stationary AR(1) state equations[54] with an autoregressive parameter 0.5. The tables below summarise the bias, MSE and coverage rates for different models and also, where applicable, for different bandwidth parameters.

Bias time-varying intercept						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
$H=T^{0.40}$	0.0030	0.0001	0.0011	-0.0075	-0.0012	-0.0112
$H=T^{0.45}$	0.0032	0.0007				
$H=T^{0.50}$	0.0034	0.0013				
$H=T^{0.55}$	0.0036	0.0018				
$H=T^{0.60}$	0.0038	0.0022				
T=100						
$H=T^{0.40}$	-0.0002	-0.0023	-0.0015	-0.0120	-0.0036	-0.0156
$H=T^{0.45}$	-0.0001	-0.0017				
$H=T^{0.50}$	0.0001	-0.0013				
$H=T^{0.55}$	0.0002	-0.0009				
$H=T^{0.60}$	0.0002	-0.0007				
T=500						
$H=T^{0.40}$	0.0039	0.0040	0.0035	0.0045	0.0054	0.0051
$H=T^{0.45}$	0.0039	0.0040				
$H=T^{0.50}$	0.0039	0.0040				
$H=T^{0.55}$	0.0039	0.0039				
$H=T^{0.60}$	0.0038	0.0038				
T=1000						
$H=T^{0.40}$	0.0018	0.0013	0.0014	-0.0025	0.0018	-0.0047
$H=T^{0.45}$	0.0018	0.0015				
$H=T^{0.50}$	0.0018	0.0015				
$H=T^{0.55}$	0.0017	0.0015				
$H=T^{0.60}$	0.0016	0.0015				

Table 21. Bias of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

[54] Unlike the Monte Carlo in Section 2.8, here we do not estimate the additional coefficients for the state space models, we simply use the state space approaches (both linear and non-linear) to filter the unobserved latent drifting parameters.

MSE time-varying intercept						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
H=T$^{\mathbf{0.40}}$	0.2080	0.1936	0.1488	0.3915	0.2293	0.4669
H=T$^{\mathbf{0.45}}$	0.1950	0.1854				
H=T$^{\mathbf{0.50}}$	0.1877	0.1816				
H=T$^{\mathbf{0.55}}$	0.1858	0.1820				
H=T$^{\mathbf{0.60}}$	0.1887	0.1864				
T=100						
H=T$^{\mathbf{0.40}}$	0.1525	0.1449	0.1083	0.3879	0.1781	0.4705
H=T$^{\mathbf{0.45}}$	0.1411	0.1367				
H=T$^{\mathbf{0.50}}$	0.1358	0.1334				
H=T$^{\mathbf{0.55}}$	0.1358	0.1346				
H=T$^{\mathbf{0.60}}$	0.1408	0.1404				
T=500						
H=T$^{\mathbf{0.40}}$	0.0815	0.0790	0.0514	0.3773	0.0904	0.4529
H=T$^{\mathbf{0.45}}$	0.0728	0.0717				
H=T$^{\mathbf{0.50}}$	0.0700	0.0696				
H=T$^{\mathbf{0.55}}$	0.0722	0.0722				
H=T$^{\mathbf{0.60}}$	0.0795	0.0797				
T=1000						
H=T$^{\mathbf{0.40}}$	0.0578	0.0566	0.0352	0.3790	0.0650	0.4565
H=T$^{\mathbf{0.45}}$	0.0511	0.0507				
H=T$^{\mathbf{0.50}}$	0.0496	0.0495				
H=T$^{\mathbf{0.55}}$	0.0529	0.0530				
H=T$^{\mathbf{0.60}}$	0.0607	0.0609				

Table 22. MSEs of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

Coverage rates time-varying intercept						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
H=T$^{\mathbf{0.40}}$	0.8730	0.9100	0.9480	0.9900	0.9560	0.9760
H=T$^{\mathbf{0.45}}$	0.8700	0.8970				
H=T$^{\mathbf{0.50}}$	0.8520	0.8770				
H=T$^{\mathbf{0.55}}$	0.8290	0.8480				
H=T$^{\mathbf{0.60}}$	0.8030	0.8180				
T=100						
H=T$^{\mathbf{0.40}}$	0.8960	0.9270	0.9480	0.9910	0.9480	0.9760
H=T$^{\mathbf{0.45}}$	0.8780	0.9080				
H=T$^{\mathbf{0.50}}$	0.8470	0.8710				
H=T$^{\mathbf{0.55}}$	0.8180	0.8310				
H=T$^{\mathbf{0.60}}$	0.7720	0.7740				
T=500						
H=T$^{\mathbf{0.40}}$	0.9170	0.9310	0.9420	0.9840	0.9360	0.9810
H=T$^{\mathbf{0.45}}$	0.8880	0.9050				
H=T$^{\mathbf{0.50}}$	0.8420	0.8580				
H=T$^{\mathbf{0.55}}$	0.7830	0.7960				
H=T$^{\mathbf{0.60}}$	0.6960	0.6960				
T=1000						
H=T$^{\mathbf{0.40}}$	0.9140	0.9240	0.9460	0.9910	0.9510	0.9830
H=T$^{\mathbf{0.45}}$	0.8940	0.8980				
H=T$^{\mathbf{0.50}}$	0.8580	0.8720				
H=T$^{\mathbf{0.55}}$	0.7870	0.7890				
H=T$^{\mathbf{0.60}}$	0.6780	0.6830				

Table 23. Coverage rates of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

Bias time-varying log volatility						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
$H=T^{0.40}$	-0.1471	0.0924				
$H=T^{0.45}$	-0.0795	0.1106				
$H=T^{0.50}$	-0.0198	0.1303	-0.0278	-0.2252	-0.0039	-0.1475
$H=T^{0.55}$	0.0341	0.1517				
$H=T^{0.60}$	0.0835	0.1749				
T=100						
$H=T^{0.40}$	-0.1121	0.0856				
$H=T^{0.45}$	-0.0541	0.0995				
$H=T^{0.50}$	-0.0028	0.1156	-0.0266	-0.2395	-0.0115	-0.1513
$H=T^{0.55}$	0.0441	0.1344				
$H=T^{0.60}$	0.0882	0.1564				
T=500						
$H=T^{0.40}$	-0.0455	0.0722				
$H=T^{0.45}$	-0.0061	0.0796				
$H=T^{0.50}$	0.0288	0.0905	-0.0059	-0.2336	0.0024	-0.1379
$H=T^{0.55}$	0.0618	0.1055				
$H=T^{0.60}$	0.0951	0.1256				
T=1000						
$H=T^{0.40}$	-0.0378	0.0539				
$H=T^{0.45}$	-0.0051	0.0596				
$H=T^{0.50}$	0.0240	0.0691	-0.0080	-0.2464	-0.0011	-0.1484
$H=T^{0.55}$	0.0523	0.0831				
$H=T^{0.60}$	0.0816	0.1024				

Table 24. Bias log volatility of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

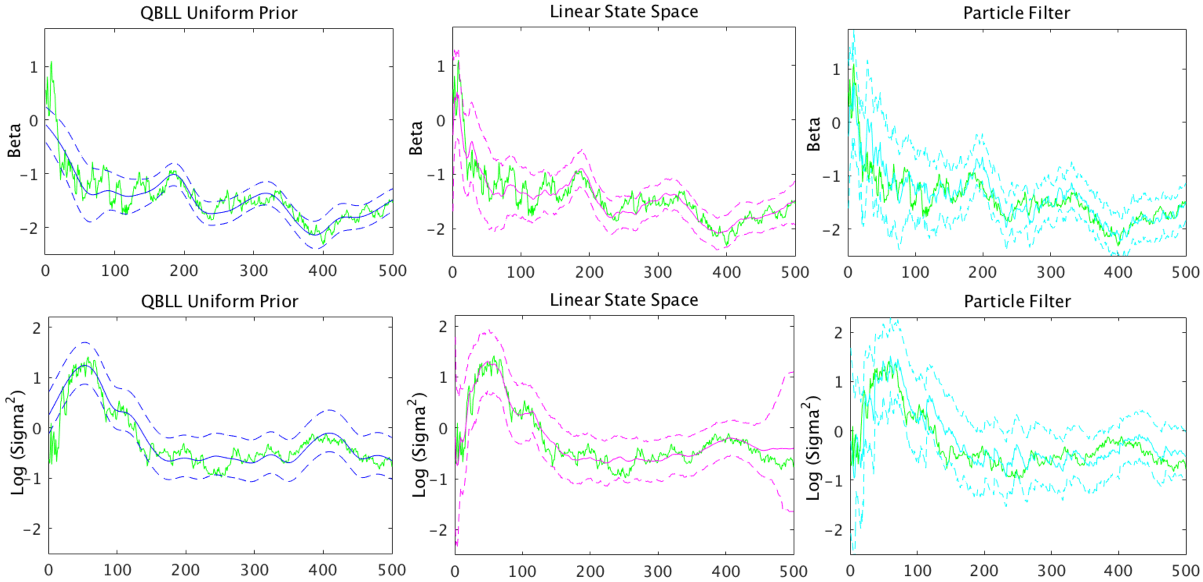
MSE time-varying log volatility						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
$H=T^{0.40}$	0.3241	0.3266				
$H=T^{0.45}$	0.2853	0.3102				
$H=T^{0.50}$	0.2659	0.3019	0.2293	0.6558	0.3668	0.7482
$H=T^{0.55}$	0.2618	0.3019				
$H=T^{0.60}$	0.2700	0.3098				
T=100						
$H=T^{0.40}$	0.2256	0.2388				
$H=T^{0.45}$	0.1979	0.2223				
$H=T^{0.50}$	0.1865	0.2151	0.1569	0.6262	0.2555	0.7205
$H=T^{0.55}$	0.1879	0.2168				
$H=T^{0.60}$	0.1999	0.2270				
T=500						
$H=T^{0.40}$	0.1062	0.1258				
$H=T^{0.45}$	0.0947	0.1131				
$H=T^{0.50}$	0.0942	0.1101	0.0687	0.6289	0.1240	0.7362
$H=T^{0.55}$	0.1023	0.1158				
$H=T^{0.60}$	0.1185	0.1298				
T=1000						
$H=T^{0.40}$	0.0771	0.0901				
$H=T^{0.45}$	0.0677	0.0791				
$H=T^{0.50}$	0.0673	0.0769	0.0478	0.6243	0.0891	0.7214
$H=T^{0.55}$	0.0747	0.0827				
$H=T^{0.60}$	0.0899	0.0965				

Table 25. MSEs log volatility of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

Coverage rates time-varying log volatility						
Bandwidth	QBLL Uniform Prior	QBLL Normal Gamma Prior	Linear SS	Linear SS Misspecification	Particle Filter	Particle Filter Misspecification
T=50						
H=T$^{\mathbf{0.40}}$	0.9400	0.8760				
H=T$^{\mathbf{0.45}}$	0.9320	0.8590				
H=T$^{\mathbf{0.50}}$	0.9150	0.8490	0.9570	0.9880	0.9490	0.9770
H=T$^{\mathbf{0.55}}$	0.8700	0.8250				
H=T$^{\mathbf{0.60}}$	0.8190	0.7790				
T=100						
H=T$^{\mathbf{0.40}}$	0.9440	0.8860				
H=T$^{\mathbf{0.45}}$	0.9350	0.8840				
H=T$^{\mathbf{0.50}}$	0.9200	0.8730	0.9510	0.9920	0.9490	0.9910
H=T$^{\mathbf{0.55}}$	0.8880	0.8470				
H=T$^{\mathbf{0.60}}$	0.8400	0.7950				
T=500						
H=T$^{\mathbf{0.40}}$	0.9540	0.9130				
H=T$^{\mathbf{0.45}}$	0.9500	0.9050				
H=T$^{\mathbf{0.50}}$	0.9220	0.8870	0.9410	0.9890	0.9460	0.9840
H=T$^{\mathbf{0.55}}$	0.8770	0.8450				
H=T$^{\mathbf{0.60}}$	0.8050	0.7700				
T=1000						
H=T$^{\mathbf{0.40}}$	0.9640	0.9310				
H=T$^{\mathbf{0.45}}$	0.9550	0.9200				
H=T$^{\mathbf{0.50}}$	0.9230	0.8900	0.9450	0.9850	0.9400	0.9760
H=T$^{\mathbf{0.55}}$	0.8690	0.8450				
H=T$^{\mathbf{0.60}}$	0.7800	0.7570				

Table 26. Coverage rates log volatility of models based on DGP in equation (89) for 50, 100, 500 and 1000 observations respectively.

Figure 44: Typical realisation of the time varying parameters and volatilities



In addition, the figures below illustrate what a typical realisation of the time varying parameter looks like, when the data are a linear regression model with a time varying intercept, generated by a non-linear smooth transition autoregressive model (STAR) model:

$$y_t = \beta_t + \sigma u_t$$

$$\beta_t = \beta_{t-1}(1 - \exp(-\gamma \beta_{t-1}^2)) + \delta \omega_t$$

$$\omega_t, \ u_t \sim NID(0, 1)$$

$$\gamma = 5, \ \sigma = 1, \ \delta = 0.5.$$

Figure 45: Typical realisation of the time varying parameters and volatilities

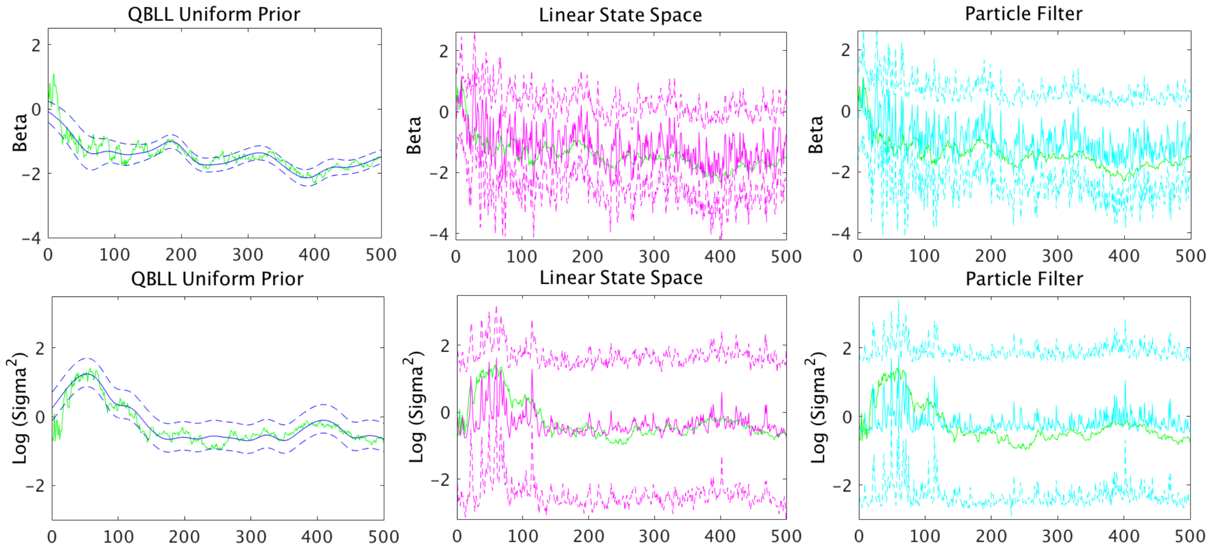
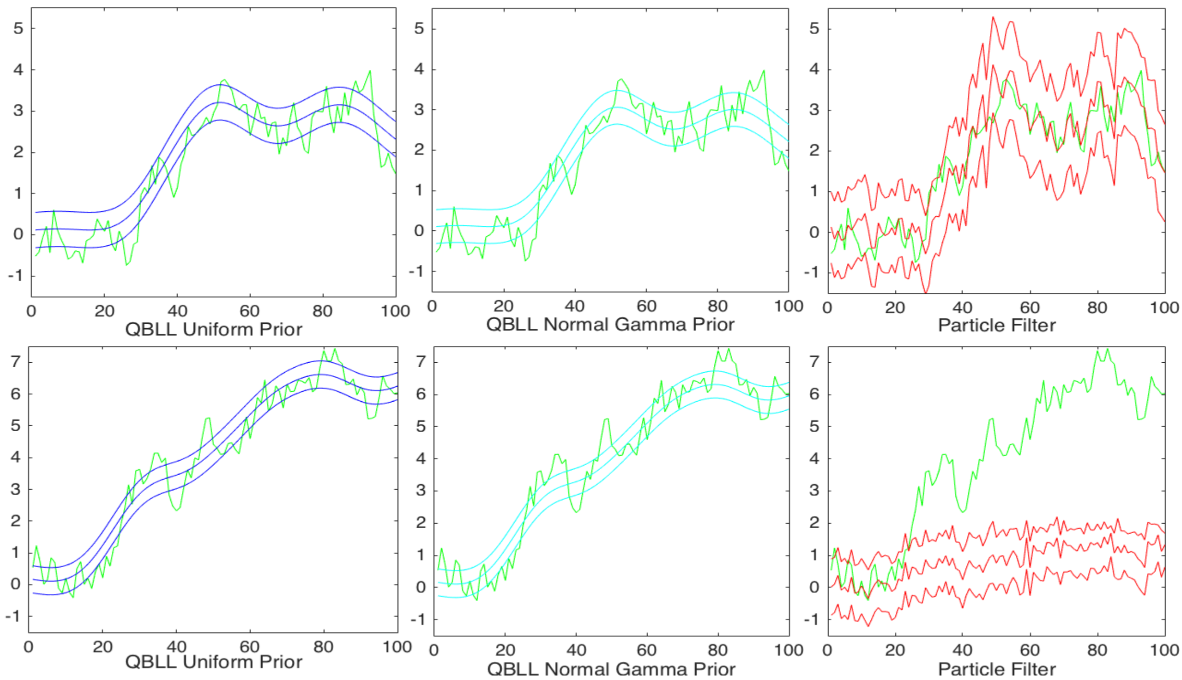


Figure 46: Typical realisation of the time varying parameters for the STAR model



7.2 Model and Data Description and Additional Results for Chapter 3

7.2.1 The Smets and Wouters (2007) Model

The resource constraint is given by:

$$y_t = \underbrace{(1 - g_y - i_y)}_{\substack{\text{steady state} \\ \text{consumption-output ratio}}} c_t + \underbrace{((\gamma - 1 - \delta)k_y)}_{\substack{\text{steady state} \\ \text{investement-output ratio}}} i_t + (R_*^k k_y) z_t + \varepsilon_t^g.$$

Output, y_t, is absorbed by consumption c_t, investment i_t, capital utilization z_t and government spending ε_t^g. g_y, i_y and k_y are steady state government-output, investment-output and capital-output ratios respectively and R_*^k is the steady state rental rate of capital. γ is the steady state growth rate of output, used to detrend all non-stationary variables in the model and δ is the depreciation rate of capital. Exogenous government spending follows an AR(1) stochastic process with an autoregressive coefficient ρ_g and an iid-Normal error term η_t^g with variance σ_g^2 :

$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \eta_t^g + \rho_{ga} \eta_t^a$$

where η_t^a is the iid-Normal shock to TFP and is motivated by Smets and Wouters (2007) as the model at hand is a closed economy, with ε_t^g also including data on exports/imports, which could depend on domestic productivity η_t^a.

The Euler equation for consumption is:

$$c_t = \frac{(\lambda/\gamma)}{(1+\lambda/\gamma)} c_{t-1} + \frac{1}{(1+\lambda/\gamma)} \mathbb{E}_t c_{t+1} + \frac{(\sigma_c - 1) W_*^h L_*/C_*}{\sigma_c (1+\lambda/\gamma)} \mathbb{E}_t (l_t - l_{t+1}) - \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c} (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b)$$

and implies that consumption c_t is a weighted average between past consumption c_{t-1} and expected future consumption $E_t c_{t+1}$. It also depends on expected growth in the hours worked $E_t(l_t - l_{t+1})$ and ex-ante real interest rate $r_t - E_t \pi_{t+1}$ and a risk premium shock ε_t^b representing a wedge between the instrument controlled by the central bank and the rate of return on assets faced by households. It follows an AR(1) stochastic process with an autoregressive coefficient ρ_b and an iid-Normal error term η_t^b with variance σ_b^2:

$$\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \eta_t^b.$$

In the absence of habit formation, $\lambda = 0$, the first term drops out and the Euler equation becomes entirely forward-looking. When the elasticity of intertemporal substitution, $\sigma_c = 1$, the household is facing log utility in consumption and the labour supply term drops out.

The investment Euler equation is:

$$i_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}} i_{t-1} + (1 - \frac{1}{1+\beta\gamma^{1-\sigma_c}})\mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi} q_t + \varepsilon_t^i$$

implying that current investment i_t is a weighted average of past investment i_{t-1} and expected future investment $E_t i_{t+1}$. It also depends on the real value of capital q_t and an investment-specific technology disturbance term ε_t^i that captures the relative efficiency of investment expenditure and follows an AR(1) stochastic process with an autoregressive coefficient ρ_i and an iid-Normal error term η_t^i with variance σ_i^2 :

$$\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \eta_t^i.$$

φ is the steady state elasticity of capital adjustment cost and β is the household's discount factor.

The arbitrage condition between the return to capital and the riskless rate is given by:

$$q_t = \frac{(1-\delta)}{R_*^k + (1-\delta)} \mathbb{E}_t q_{t+1} + (1 - \frac{(1-\delta)}{R_*^k + (1-\delta)})\mathbb{E}_t r_{t+1}^k - (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b)$$

where the current capital value q_t is a weighted average of expected future value $E_t q_{t+1}$ and expected real rental rate of capital $E_t r_{t+1}^k$ and depends also negatively on the ex-ante real interest rate $r_t - E_t \pi_{t+1}$ and the risk-premium disturbance ε_t^b. The aggregate production function is:

$$y_t = \phi(\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a)$$

with output being produced with standard factors of production, capital k_t^s, labour l_t and technology ε_t^a, assumed to follow an AR(1) stochastic process with an autoregressive coefficient ρ_a and an iid-Normal error term η_t^a with variance σ_a^2:

$$\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a.$$

The parameter ϕ measures one plus the fixed costs in production and α is the capital share after mark-ups and fixed costs. Since the capital installed today takes a period to become effective, current capital used in production k_t^s is a sum of capital installed the previous period k_{t-1} and the degree of capital utilization z_t:

$$k_t^s = k_{t-1} + z_t.$$

The degree of capital utilization z_t itself depends positively in the rental rate of capital r_t^k and ψ is a function of the elasticity of capital utilization adjustment cost function, normalised to take valued between zero and one:

$$z_t = \frac{1-\psi}{\psi} r_t^k.$$

The capital accumulation equation is given by:

$$k_t = \frac{1-\delta}{\gamma} k_{t-1} + (1 - \frac{1-\delta}{\gamma}) i_t + (1 - \frac{1-\delta}{\gamma})((1 + \beta\gamma^{1-\sigma_c})\gamma^2 \varphi)\varepsilon_t^i$$

where installed capital k_t is a function of previously installed capital k_{t-1}, investment flow i_t and the investment-specific disturbance ε_t^i.

The price mark-up μ_t^p in the monopolistic goods market is given by the difference between marginal product of labour mpl_t, which depends on TFP and the capital-labour ratio, and the real wage w_t:

$$\mu_t^p = \underbrace{\alpha(k_t^s - l_t) + \varepsilon_t^a}_{mpl_t} - w_t.$$

The Phillips curve is given by:

$$\pi_t = \frac{\iota_p}{1 + \beta\gamma^{(1-\sigma_c)}\iota_p} \pi_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1 + \beta\gamma^{(1-\sigma_c)}\iota_p} \mathbb{E}_t \pi_{t+1} - \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}\iota_p} \left\{ \frac{(1 - \beta\gamma^{(1-\sigma_c)}\xi_p)(1 - \xi_p)}{\xi_p((\phi - 1)\varepsilon_p + 1)} \right\} \mu_t^p + \varepsilon_t^p$$

and implies that current level of inflation π_t is a function of past inflation π_{t-1} and expected future inflation $E_t \pi_{t+1}$, price mark-up μ_t^p and a price mark-up shock ε_t^p. Without degree of indexation, $\iota_p = 0$, the expression reduces to purely forward-looking Phillips curve. ξ_p is the Calvo price stickiness, ε_p is the Kimball aggregator in the goods market that measures the degree of strategic interaction between price-setters and $\phi - 1$ is the steady state price mark-up, which depends on the fixed cost parameter ϕ. The price mark-up error term ε_t^p follows an ARMA(1,1) process with an autoregressive coefficient ρ_p, a moving average coefficient μ_p and an iid-Normal error term η_t^p with variance σ_p^2, motivated by the desire to capture more of the dynamics in the data on inflation fluctuations:

$$\varepsilon_t^p = \rho_p \, \varepsilon_{t-1}^p + \eta_t^p + \mu_p \eta_{t-1}^p.$$

Rental rate of capital is a function of the capital-labour ratio $(k_t - l_t)$ and the real wage w_t:

$$r_t^k = -(k_t - l_t) + w_t.$$

The labour market is characterised by similar conditions to the goods market. In particular, there is a wage mark-up equation:

$$\mu_t^w = w_t - \underbrace{(\sigma_l l_t + \frac{1}{1-\lambda/\gamma}(c_t - \lambda/\gamma c_{t-1}))}_{mrs_t}$$

where the wage mark-up μ_t^w is the difference between the real wage w_t and the marginal rate of substitution between working and consuming, mrs_t, that is the disutility of work, with σ_l capturing the elasticity of labour supply with respect to the wage. The corresponding wage equation is given by:

$$\begin{aligned} w_t = & \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} w_{t-1} + (1 - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}})(\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}) - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} \pi_t \\ & + \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}} \pi_{t-1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} \left\{ \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{\xi_w((\phi_w - 1)\varepsilon_w + 1)} \right\} \mu_t^w + \varepsilon_t^w. \end{aligned}$$

The real wage w_t is a weighted average between past wage w_{t-1} and expected future real wage $(E_t w_{t+1} + E_t \pi_{t+1})$, depends on wage mark-up, current inflation π_t, wage mark-up shock ε_t^w and partially indexed to past inflation π_{t-1}. Similarly to the goods market, ι_w captures the degree of indexation, ξ_w is the Calvo wage stickiness, ε_w is the Kimball aggregator in the labour market and $\phi_w - 1$ is the steady state wage mark-up. Finally, the wage disturbance also follows an ARMA(1,1) process with an autoregressive coefficient ρ_w, a moving average coefficient μ_w and an iid-Normal error term η_t^w with variance σ_w^2, with the MA term added as explained by Smets and Wouters (2007) to capture more of the high frequency wage fluctuations observed in the data:

$$\varepsilon_t^w = \rho_w \, \varepsilon_{t-1}^w + \eta_t^w + \mu_w \eta_{t-1}^w.$$

The central bank in the model follows a nominal interest rate rule of the form:

$$r_t = \rho r_{t-1} + (1-\rho)\left\{r_\pi \pi_t + r_y(y_t - y_t^p)\right\} + r_{\Delta y}\left((y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)\right) + \varepsilon_t^r$$

by gradually adjusting the policy rate r_t in response to fluctuations in inflation π_t, output gap

$(y_t - y_t^p)$ and output gap growth $(y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)$. The policy parameters ρ, r_π, r_y and $r_{\Delta y}$ capture the degree of interest rate smoothing, the level of inflation and output gap targetting and the short-run feedback from output gap change respectively. The monetary policy shock ε_t^r follow an AR(1) stochastic process with an autoregressive coefficient ρ_r and an iid-Normal error term η_t^r with variance σ_r^2:

$$\varepsilon_t^r = \rho_r \; \varepsilon_{t-1}^r + \eta_t^r.$$

The measurement equation takes the form:

$$X_t = \begin{bmatrix} 100 \times \Delta \log GDP_t \\ 100 \times \Delta \log C_t \\ 100 \times \Delta \log I_t \\ 100 \times \Delta \log W_t \\ 100 \times \log H_t \\ 100 \times \Delta \log P_t \\ FFR_t \end{bmatrix} = \begin{bmatrix} \overline{\gamma} \\ \overline{\gamma} \\ \overline{\gamma} \\ \overline{\gamma} \\ \overline{l} \\ \overline{\pi} \\ \overline{r} \end{bmatrix} + \begin{bmatrix} y_t - y_{t-1} \\ c_t - c_{t-1} \\ i_t - i_{t-1} \\ w_t - w_{t-1} \\ l_t \\ \pi_t \\ r_t \end{bmatrix}$$

7.2.2 Additional Time Varying IRFs

Figure 47: IRFs to 1 unit investment technology shock

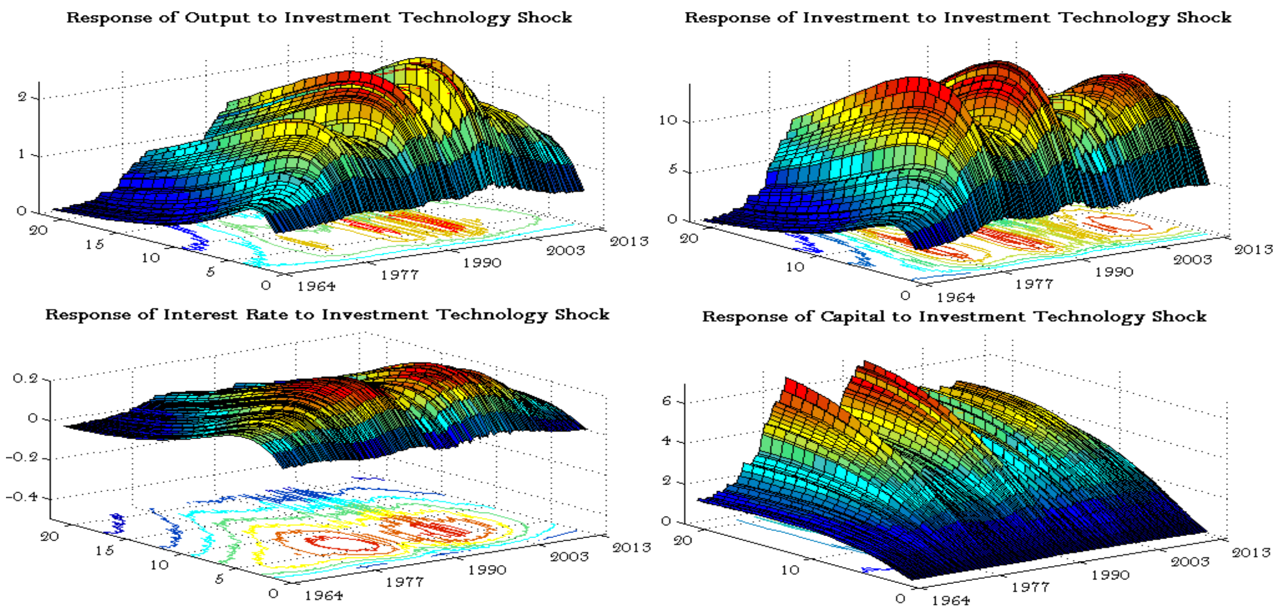


Figure 48: IRFs to 1 st. dev. investment technology shock

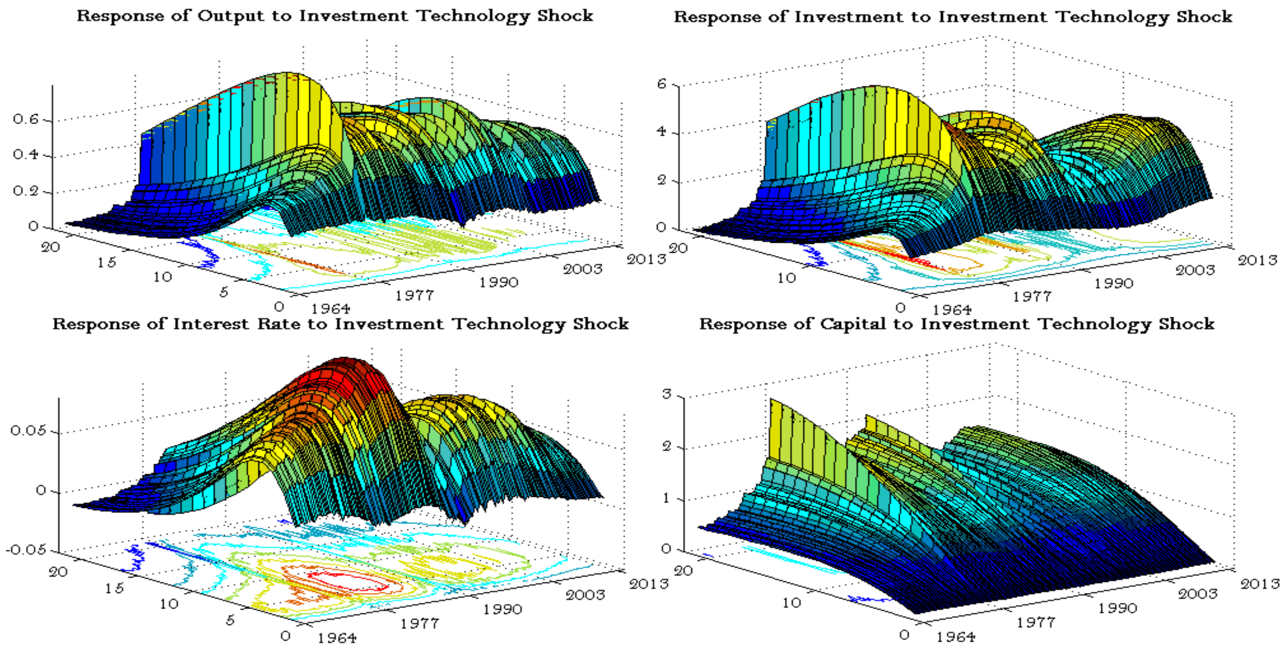


Figure 49: IRFs to 1 unit government spending shock

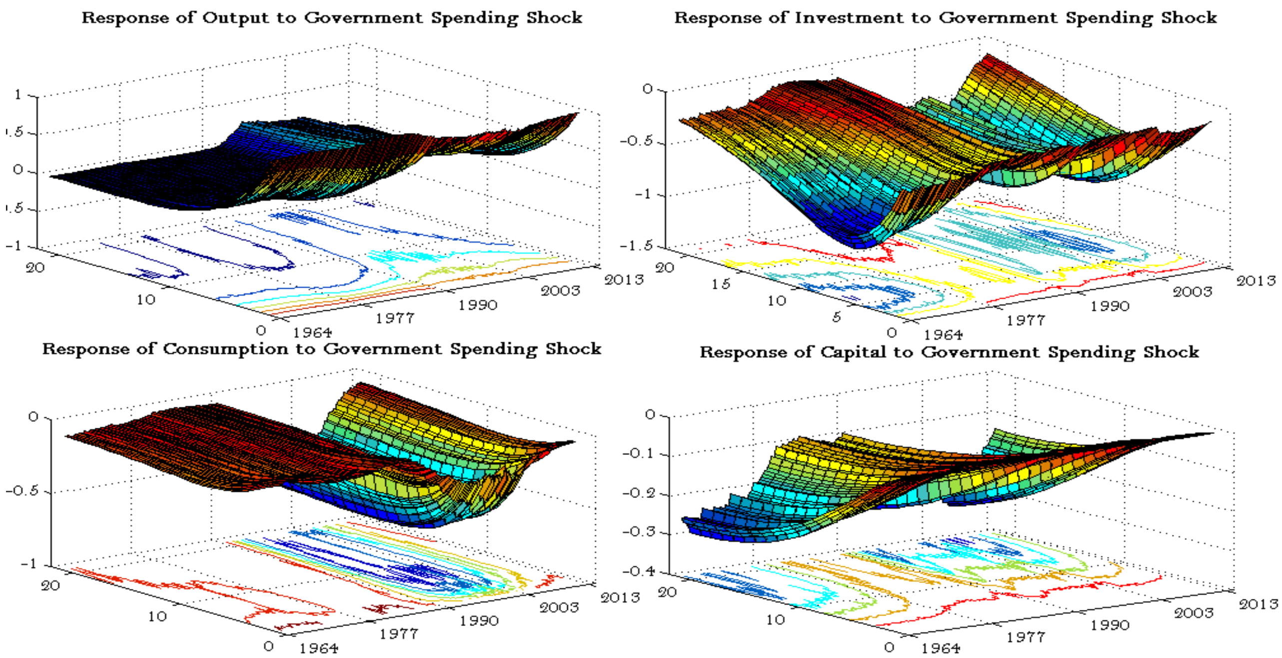
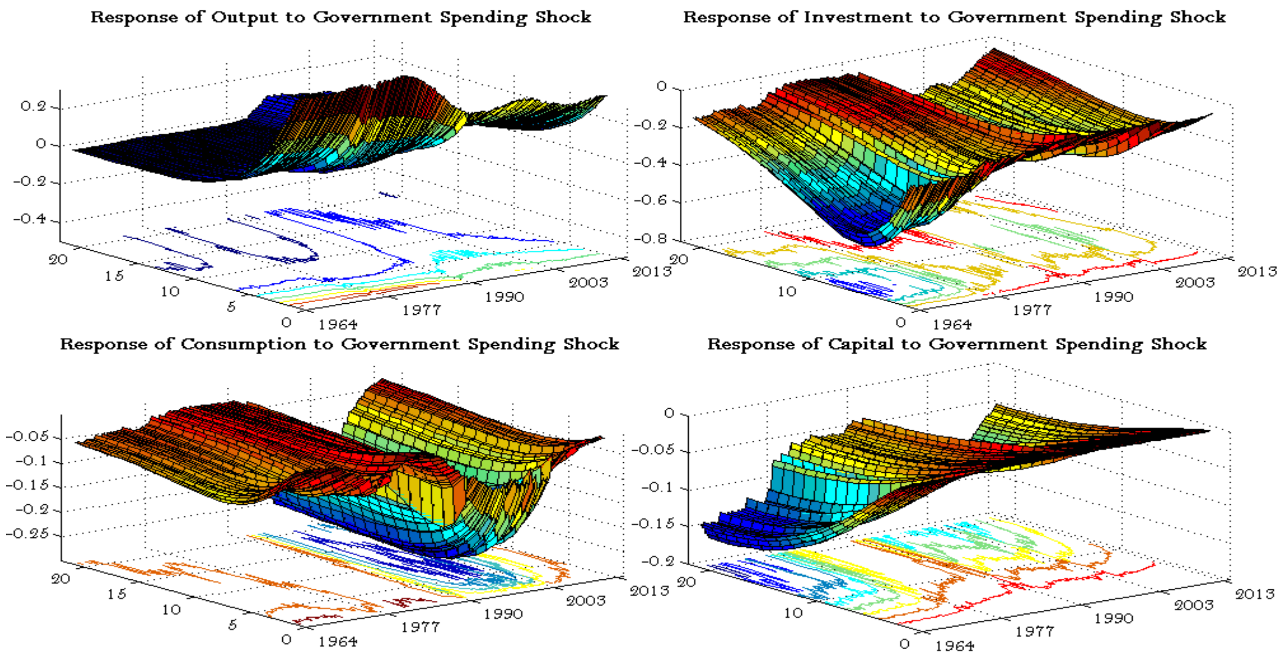


Figure 50: IRFs to 1 st. dev. government spending shock



7.3 Model and Data Descriptions and Additional Results for Chapter 4

7.3.1 The Smets and Wouters (2007) model with financial frictions

The model we use is a Smets and Wouters (2007) model with a deterministic trend, modified to include a financial friction block, as in Bernanke et al. (1999). We refer the reader to the original paper, Smets and Wouters (2007), for discussion and derivation of the model's equation and for completeness, we list here the linearised equations. See also the Technical Appendix in Smets and Wouters (2007) available at: http://www.aeaweb.org/aer/data/june07/20041254_app.pdf. For expressions of the FF block parameters and steady states, see:

http://sites.sas.upenn.edu/schorf/files/hb_forecasting_appendix.pdf, pp 37-39.

- The resource constraint in the model is given by equation,

$$y_t = (1 - g_y - i_y)c_t + ((\gamma - 1 - \delta)k_y)i_t + (R_*^k k_y)z_t + \varepsilon_t^g,$$

- the consumption Euler equation,

$$c_t = \frac{(\lambda/\gamma)}{(1+\lambda/\gamma)}c_{t-1} + \frac{1}{(1+\lambda/\gamma)}\mathbb{E}_t c_{t+1} + \frac{(\sigma_c - 1)W_*^h L_*/C_*}{\sigma_c(1+\lambda/\gamma)}\mathbb{E}_t(l_t - l_{t+1}) - \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c}(r_t - \mathbb{E}_t\pi_{t+1} + \varepsilon_t^b),$$

- the investment Euler equation,

$$i_t = \frac{1}{1+\beta\gamma^{1-\sigma_c}} i_{t-1} + (1 - \frac{1}{1+\beta\gamma^{1-\sigma_c}})\mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi} q_t + \varepsilon_t^i,$$

- the aggregate production function, $y_t = \phi(\alpha k_t^s + (1-\alpha)l_t + \varepsilon_t^a)$,

- the relation between effectively rented capital and capital, $k_t^s = k_{t-1} + z_t$,

- the degree of capital utilization, $z_t = \frac{1-\psi}{\psi} r_t^k$,

- the capital accumulation equation, $k_t = \frac{1-\delta}{\gamma} k_{t-1} + (1 - \frac{1-\delta}{\gamma}) i_t + (1 - \frac{1-\delta}{\gamma})((1+\beta\gamma^{1-\sigma_c})\gamma^2\varphi)\varepsilon_t^i$,

- the price mark-up, $\mu_t^p = \alpha(k_t^s - l_t) + \varepsilon_t^a - w_t$,

- the new Keynesian Phillips curve,

$$\pi_t = \frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\pi_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\mathbb{E}_t\pi_{t+1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}\iota_p}\left\{\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{\xi_p((\phi-1)\varepsilon_p+1)}\right\}\mu_t^p + \varepsilon_t^p,$$

- the rental rate of capital, $r_t^k = -(k_t - l_t) + w_t$,

- the wage mark-up, $\mu_t^w = w_t - (\sigma_l l_t + \frac{1}{1-\lambda/\gamma}(c_t - \lambda/\gamma c_{t-1}))$,

- the wage equation,

$$\begin{aligned} w_t \quad = \quad & \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} w_{t-1} + (1 - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}})(\mathbb{E}_t w_{t+1} + \mathbb{E}_t\pi_{t+1}) - \frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\pi_t \\ & + \frac{\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\pi_{t-1} - \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\left\{\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{\xi_w((\phi_w-1)\varepsilon_w+1)}\right\}\mu_t^w + \varepsilon_t^w, \end{aligned}$$

- the Taylor Rule,

$$r_t = \rho r_{t-1} + (1-\rho)\{r_\pi \pi_t + r_y(y_t - y_t^p)\} + r_{\Delta y}\left((y_t - y_t^p) - (y_{t-1} - y_{t-1}^p)\right) + \varepsilon_t^r.$$

The financial friction block

- The corporate spread is defined as

$$\mathbb{E}_t\left[\tilde{R}_{t+1}^k - r_t\right] = \frac{(1-\lambda/\gamma)}{(1+\lambda/\gamma)\sigma_c}\varepsilon_t^b + \varsigma_{sp,b}(q_t + \overline{k}_t - n_t) + \varepsilon_t^\omega.$$

- The arbitrage condition between the return to capital and the riskless rate in Smets and Wouters (2007) is now replaced by

$$\tilde{R}_t^k - \pi_t = \frac{r_*^k}{r_*^k + (1-\delta)} r_t^k + \frac{(1-\delta)}{r_*^k + (1-\delta)} q_t - q_{t-1}.$$

- Finally, the entrepreneurs' net worth evolution is defined as

$$n_t = \varsigma_{n,R^K}(\tilde{R}_t^k - \pi_t) - \varsigma_{n,R}(r_{t-1} - \pi_t) + \varsigma_{n,q}(q_{t-1} + \bar{k}_{t-1}) + \varsigma_{n,n} n_{t-1} - \frac{\varsigma_{n,\omega}}{\varsigma_{sp,\omega}} \varepsilon_{t-1}^{\omega}.$$

Stochastic processes of exogenous shocks

- Exogenous government spending spending is defined as $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \sigma_g \eta_t^g + \rho_{ga} \sigma_z \eta_t^z$,
- TFP shock, $\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \sigma_a \eta_t^a$,
- risk premium shock, $\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \sigma_b \eta_t^b$,
- investment-specific technology shock, $\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \sigma_i \eta_t^i$,
- monetary policy shock, $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \eta_t^r$,
- price mark-up shock, $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \sigma_p \eta_t^p + \mu_p \sigma_p \eta_{t-1}^p$,
- wage mark-up shock, $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \sigma_w \eta_t^w + \mu_w \sigma_w \eta_{t-1}^w$,
- financial friction shock, $\varepsilon_t^\omega = \rho_\omega \varepsilon_{t-1}^\omega + \sigma_\omega \eta_t^\omega$.

7.3.2 Measurement equation, data description and transformations

Measurement equation

$$Y_t = \begin{bmatrix} \text{Output Growth}_t \\ \text{Consumption Growth}_t \\ \text{Investment Growth}_t \\ \text{Wage Growth}_t \\ \text{Hours Worked}_t \\ \text{Inflation}_t \\ \text{Policy Rate}_t \\ \text{Spread}_t \end{bmatrix} = \begin{bmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \\ SP^* \end{bmatrix} + \begin{bmatrix} y_t - y_{t-1} \\ c_t - c_{t-1} \\ i_t - i_{t-1} \\ w_t - w_{t-1} \\ l_t \\ \pi_t \\ r_t \\ 100 * \mathbb{E}_t(\tilde{R}_{t+1}^k - r_t) \end{bmatrix}.$$

Data description

Data Description	
Variable	Source
GDP, Total, Constant Prices, AR, SA, USD, 2009 chnd prices	U.S. Bureau of Economic Analysis
PCE, Total, Constant Prices, AR, SA, USD, 2009 chnd prices	U.S. Bureau of Economic Analysis
Private Fixed Investment, Total, Current Prices, AR, SA, USD	U.S. Bureau of Economic Analysis
Consumer price index, AR, SA, Index, 2005=100	U.S. Bureau of Economic Analysis
Real hourly compensation, nonfarm business, index, SA, Index, 2009=100	U.S. Bureau of Labor Statistics
Hours worked per employee, AR	U.S. Bureau of Labor Statistics
Employment, all persons (ages 15 and over), SA	U.S. Bureau of Labor Statistics
Popultaion Total (Estimates Used in National Accounts)	U.S. Bureau of Economic Analysis
Federal Funds Rate (Monthly Average)	Federal Reserve, U.S.
Moody's Baa-Rated Long-Term, Yield, Average, USD	Reuters
Constant Maturity Yields, 10 Year, USD	Federal Reserve, U.S.

Table 27. Data Description for DSGE with FF in Chapter 3.

Data transformations:

$$\text{Output Growth}_t = 100 * \Delta \ln(GDP_t/POP_t)$$

$$\text{Consumption Growth}_t = 100 * \Delta \ln(CON_t/POP_t)$$

$$\text{Investment Growth}_t = 100 * \Delta \ln((INV_t/CPI_t)/POP_t)$$

$$\text{Wage Growth}_t = 100 * \Delta \ln(WAGE_t)$$

$$\text{Hours Worked}_t = 100 * \ln\left[((EMPL_t * HOURS_t)/POP_t) - \overline{(EMPL_t * HOURS_t)/POP_t)}\right]$$

$$\text{Inflation}_t = 100 * \Delta \ln(CPI_t)$$

$$\text{Policy Rate}_t = 1/4 * FFR_t$$

$$\text{Spread}_t = 1/4 * (BAA\_Yield_t - 10YTreasury\_Yield_t)$$

Priors

Table 28: Prior distributions for the structural parameters[55].

	Parameter Name	Prior Distribution		
		Distribution	Mean	St. Dev.
φ	Elasticity of Capital Adjustment Cost Function	Normal	4	1.5
σ_c	Elasticity of Intertemporal Substitution	Normal	1.5	0.37
λ	External Habit Formation	Beta	0.7	0.1
ξ_w	Calvo Probability in Labour Markets	Beta	0.5	0.1
σ_l	Elasticity of Labour Supply to Real Wage	Normal	2	0.75
ξ_p	Calvo Probability in Goods Markets	Beta	0.5	0.1
ι_w	Degree of Wage Indexation	Beta	0.5	0.15
ι_w	Degree of Price Indexation	Beta	0.5	0.15
ψ	Normalized Elasticity of Capital	Beta	0.5	0.2
Φ	Fixed Costs of Intermediate Goods Producers	Normal	1.25	0.12
r_π	Inflation Coefficient in the Taylor Rule	Normal	1.5	0.25
ρ	Interest Rate Smoothing Coefficient	Beta	0.75	0.1
r_y	Output Gap Coefficient in the Taylor Rule	Normal	0.12	0.05
$r_{\Delta y}$	Short-Run Feedback of Output Gap Change	Normal	0.12	0.05
$100(\beta^{-1}-1)$	Normalized Households' Discount Factor	Gamma	0.25	0.1
π^*	Steady State Inflation Rate	Gamma	0.62	0.1
l^*	Steady State Hours Worked	Normal	0	2
γ^*	Steady State Quarterly Growth Rate	Normal	0.4	0.1
α	Capital Share	Normal	0.3	0.05
SP^*	Steady State Spread	Gamma	2	0.3
$\zeta_{sp,b}$	Effect of spread on Tobin's Q, capital and networth	Beta	0.05	0.015

Table 29: Prior distributions for the parameters of the exogenous processes.

	Parameter Name	Prior Distribution		
		Distribution	Mean	St. Dev.
σ_a	St. Dev. Of TFP Shock	Inverse Gamma	0.1	2
σ_b	St. Dev. of Risk Premium Shock	Inverse Gamma	0.1	2
σ_g	St. Dev. of Exogenous Spending Shock	Inverse Gamma	0.1	2
σ_l	St. Dev. of Investment-Specific Technology Shock	Inverse Gamma	0.1	2
σ_r	St. Dev. of Monetary Policy Shock	Inverse Gamma	0.1	2
σ_p	St. Dev. of Price Mark-Up Shock	Inverse Gamma	0.1	2
σ_w	St. Dev. of Wage Mark-Up Shock	Inverse Gamma	0.1	2
σ_ω	St. Dev. of Financial Friction Shock	Inverse Gamma	0.1	2
ρ_a	Persistence Coefficient of TFP Shock	Beta	0.5	0.2
ρ_b	Persistence Coefficient of Risk Premium Shock	Beta	0.5	0.2
ρ_g	Persistence Coefficient of Spending Shock	Beta	0.5	0.2
ρ_l	Persistence Coefficient of Investment Shock	Beta	0.3	0.2
ρ_r	Persistence Coefficient of Monetary Policy Shock	Beta	0.3	0.2
ρ_p	Persistence Coefficient of Price Mark Up Shock	Beta	0.3	0.2
ρ_w	Persistence Coefficient of Wage Mark Up Shock	Beta	0.5	0.2
ρ_ω	Persistence Coefficient of Financial Friction Shock	Beta	0.5	0.2
μ_p	MA Coefficient of Price Mark Up Shock	Beta	0.5	0.2
μ_w	MA Coefficient of Wage Mark Up Shock	Beta	0.5	0.2
ρ_{ga}	Coefficient for TFP Shock in the Spending Equation	Normal	0.5	0.2

[55] We also have upon request time varying estimation results with prior standard deviations 0.1 and 0.005 for SP^* and $\zeta_{sp,b}$ respectively and with prior standard deviations 1 and 0.1 for SP^* and $\zeta_{sp,b}$ respectively. Results on the FF block remain robust to these two specifications.

Additional results

Figures 51 and 52 are quasi-posterior estimates for the additonal parameters of the DSGE with FF in Chapter 4. The posterior mean obtained by QBLL in the blue solid line, the 5% and 95% posterior quantile values are the black dotted lines, the posterior mode obtained by QBLL is the pink dash-dotted line, the posterior mean obtained by fixed Bayesian estimation is the dashed blue line, and the 5% and 95% posterior quantiles are the green dashed lines.

Figure 51: Additional QBLL Estimates

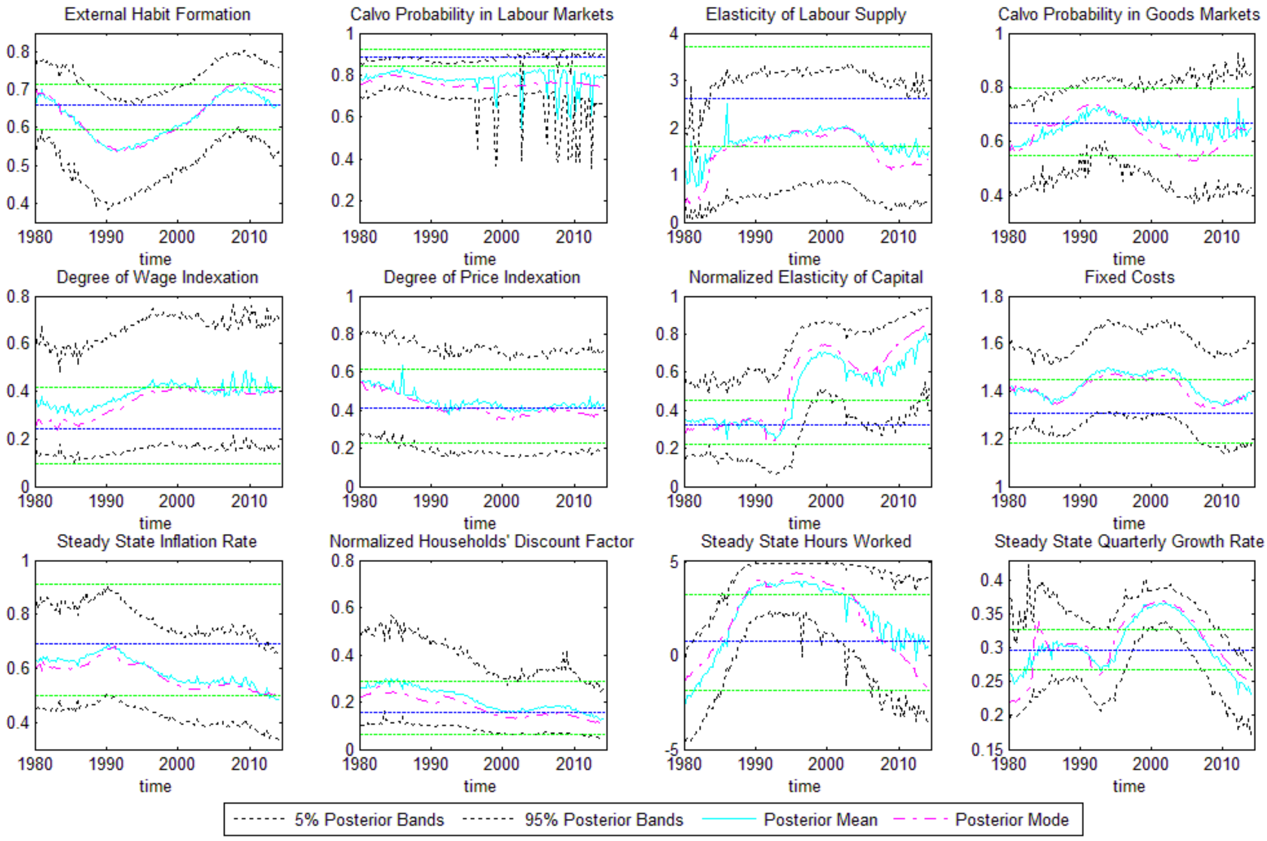
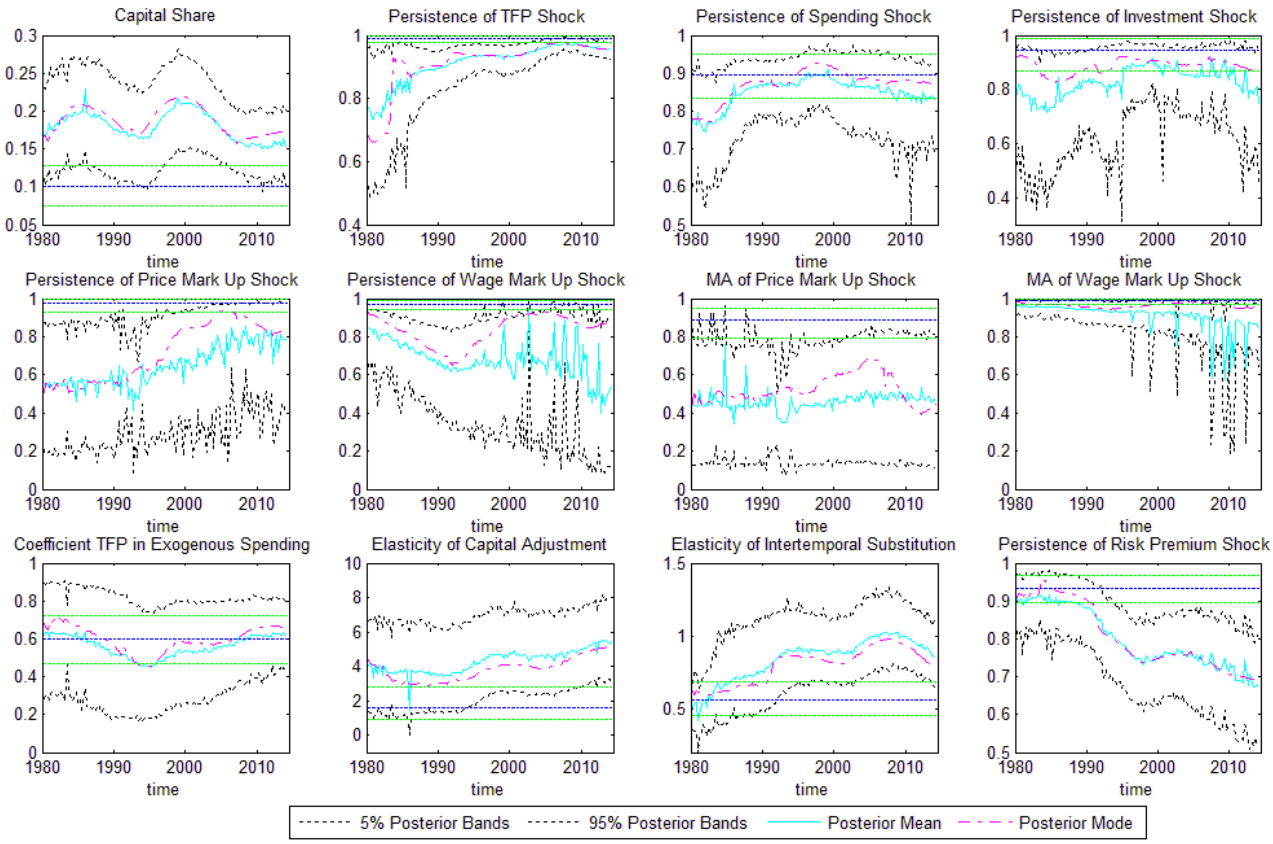


Figure 52: Additional QBLL Estimates



Additional forecasting results

	2000Q1-2006Q4				2007Q1-2012Q2			
	TV FF relative to SW				TV FF relative to SW			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Consumption Growth	1.04	1.03	1.13	1.30	1.02	0.97	1.11	1.57
Wage Growth	1.10	1.03	1.03	1.02	1.14	1.13	1.04	1.04
Hours Worked	1.01	0.869*	0.668*	0.64	0.87**	0.85*	0.85	0.84
	Fixed FF relative to SW				Fixed FF relative to SW			
Consumption Growth	1.45	1.64	1.90	1.87	1.13	1.15	1.23	1.54
Wage Growth	1.06	1.04	1.00	0.99	1.02	1.02	0.99	1.01
Hours Worked	1.01	0.98	1.02	1.22	1.00	0.94	0.87*	0.81*
	2000Q1-2006Q4				2007Q1-2012Q2			
	Density TV FF relative to SW				Density TV FF relative to SW			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Consumption Growth	-0.03	-0.04	-0.04	-0.11	-0.04	-0.06	-0.18	-0.30
Wage Growth	-0.18	0.05	0.02	0.04	0.13	0.53	0.39*	0.15
Hours Worked	0.043*	0.082*	0.210*	0.28	0.16**	0.53	2.08	2.79
	Density Fixed FF relative to SW				Density Fixed FF relative to SW			
Consumption Growth	-0.34	-0.43	-0.45	-0.43	-0.19	-0.16	-0.19	-0.26
Wage Growth	-0.15	-0.06	0.00	0.01	-0.20	0.24	-0.08	-0.02
Hours Worked	0.032*	-0.01	-0.06	-0.21	-0.12	0.11	1.76	2.88

Table 30: RMSFEs and Log Scores for additional variables. The table reports ratios of RMSFEs relative to the SW model RMSFEs and differences of log predictive scores from SW model log scores. '*', '**' and '***' indicate rejection of the null of equal performance against the one-sided alternative at 10%, 5% and 1% respectively, using Diebold and Mariano test.

	2000Q1-2006Q4				2007Q1-2012Q2			
	TV FF relative to AR(1)				TV FF relative to AR(1)			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	0.85	0.98	1.04	1.16	0.82	0.84	0.91	1.07
Consumption Growth	1.32	1.49	1.34	1.13	0.777*	0.720*	0.83	1.11
Investment Growth	0.96	0.86	0.79	0.91	1.11	1.22	1.41	1.38
Wage Growth	1.04	1.06	1.05	1.07	1.20	1.18	1.12	1.08
Hours Worked	0.92	0.77	0.56	0.49	0.711**	0.754*	0.84	0.87
Inflation	0.906*	0.90	0.876*	0.94	1.02	1.01	0.99	0.730*
Fed Funds Rate	0.666***	0.704**	0.716*	0.660*	0.742*	0.713*	0.82	1.04
	Fixed FF relative to AR(1)				Fixed FF relative to AR(1)			
Output Growth	1.02	1.25	1.45	1.51	0.84	0.81	0.82	0.93
Consumption Growth	1.83	2.37	2.25	1.62	0.86	0.86	0.92	1.09
Investment Growth	1.03	0.97	0.92	0.98	1.10	1.15	1.33	1.42
Wage Growth	1.01	1.07	1.03	1.04	1.08	1.06	1.06	1.06
Hours Worked	0.93	0.87	0.86	0.94	0.817***	0.828*	0.85	0.83
Inflation	0.94	0.96	0.98	1.05	0.97	0.94	0.94	0.86
Fed Funds Rate	0.610***	0.621**	0.704*	0.74	1.13	1.03	0.95	0.98
	2000Q1-2006Q4				2007Q1-2012Q2			
	Density TV FF relative to AR(1)				Density TV FF relative to AR(1)			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	0.00	-0.09	-0.17	-0.24	0.06	-0.14	-0.23	-0.38
Consumption Growth	-0.18	-0.32	-0.21	-0.16	0.272*	0.19	0.03	-0.41
Investment Growth	0.11	0.248**	0.352**	0.21	-0.07	-0.32	-0.80	-0.67
Wage Growth	0.321*	0.375**	0.400**	0.380*	0.07	0.27	0.398*	0.336*
Hours Worked	0.00	0.17	0.35	0.59	0.378*	0.40	3.40	5.35
Inflation	0.251*	-0.03	0.158**	0.12	0.37	0.29	-0.19	-0.09
Fed Funds Rate	0.477***	0.250*	0.05	0.24	0.10	0.14	-0.03	-0.53
	Density Fixed FF relative to AR(1)				DensityFixed FF relative to AR(1)			
Output Growth	-0.22	-0.33	-0.45	-0.48	0.02	-0.11	-0.02	-0.24
Consumption Growth	-0.48	-0.70	-0.63	-0.48	0.12	0.09	0.02	-0.36
Investment Growth	-0.08	-0.03	0.01	0.04	0.13	-0.11	-0.39	-0.56
Wage Growth	0.351**	0.274**	0.374**	0.348*	-0.27	-0.01	-0.07	0.165*
Hours Worked	-0.01	0.08	0.09	0.10	0.10	-0.03	3.09	5.43
Inflation	0.329**	0.01	0.02	-0.08	-8.12	-2.72	-1.14	0.18
Fed Funds Rate	0.02	0.05	0.13	0.30	-0.35	-0.17	0.02	0.06

Table 31: RMSFEs and Log Scores: Comparison with AR(1). The table reports ratios of RMSFEs relative to an AR(1) model RMSFEs and differences of log predictive scores from an AR(1) model log scores. '*', '**' and '***' indicate rejection of the null of equal performance against the one-sided alternative at 10%, 5% and 1% respectively, using Diebold and Mariano test.

	2000Q1-2006Q4				2007Q1-2012Q2			
	TV FF relative to TVP AR(1)				TV FF relative to TVP AR(1)			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	0.759*	0.89	0.94	1.11	0.83	0.84	0.88	0.78
Investment Growth	0.811**	0.747**	0.659*	0.772*	1.09	1.14	1.09	0.87
Inflation	1.08	1.07	1.11	1.11	0.88	1.22	1.19	0.67
Fed Funds Rate	0.604***	0.602**	0.589**	0.465*	0.80	0.76	0.84	0.96
	Fixed FF relative to TVP AR(1)				Fixed FF relative to TVP AR(1)			
Output Growth	0.91	1.14	1.31	1.44	0.84	0.80	0.79	0.68
Investment Growth	0.871*	0.84	0.77	0.83	1.07	1.08	1.03	0.89
Inflation	1.12	1.15	1.24	1.24	0.84	1.14	1.13	0.79
Fed Funds Rate	0.553***	0.531**	0.579**	0.521*	1.21	1.10	0.97	0.90
	2000Q1-2006Q4				2007Q1-2012Q2			
	Density TV FF relative to TVP AR(1)				Density TV FF relative to TVP AR(1)			
	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8
Output Growth	0.248*	0.01	-0.11	-0.20	0.12	0.10	0.02	0.00
Investment Growth	0.228*	0.390**	0.560**	0.540**	0.12	-0.18	-0.33	0.24
Inflation	-0.04	-0.14	-0.04	0.07	-2.17	-0.22	-0.45	0.07
Fed Funds Rate	0.672***	0.569**	0.44	1.010*	-0.11	-0.03	-0.14	-0.54
	Density Fixed FF relative to TVP AR(1)				DensityFixed FF relative to TVP AR(1)			
Output Growth	0.03	-0.24	-0.39	-0.44	0.07	0.13	0.23	0.14
Investment Growth	0.04	0.11	0.22	0.367*	0.31	0.03	0.07	0.36
Inflation	0.04	-0.11	-0.17	-0.13	-10.65	-3.23	-1.39	0.33
Fed Funds Rate	0.21	0.369*	0.520*	1.072**	-0.57	-0.33	-0.09	0.04

Table 32: RMSFEs and Log Scores for selected variables. The table reports ratios of RMSFEs relative to a TVP AR(1) model RMSFEs and differences of log predictive scores from a TVP AR(1) model log scores. '*', '**' and '***' indicate rejection of the null of equal performance against the one-sided alternative at 10%, 5% and 1% respectively, using Diebold and Mariano test.

For the time varying parameter (TVP) AR(1), the model is estimated in each point in time t : $\widehat{\beta}_t = (X'D_tX)^{-1}X'D_tY$ where X contains the lagged dependent variable Y and D_t is a diagonal matrix with the kernel weights of the t^{th} row of the weighting matrix in equation (4) in its main diagonal. The variance of the residuals is also time varying and computed in point t as $\widehat{\sigma}_t^2 = \varepsilon' D_t \varepsilon / tr(D_t)$. Density forecasts are then generated, using wild bootstrap and the last period values $\widehat{\beta}_T$ and $\widehat{\sigma}_T^2$.

7.3.3 Robustness checks: flat kernel

Figures 53, 54 and 55 display the additional robustness checks for Chapter 4. The posterior mode obtained by QBLL is the pink dash-dotted line, the 5% and 95% posterior quantile values are the black dotted lines, the posterior mode obtained by rolling window is the solid green line.

Figure 53: Additional Robustness Checks

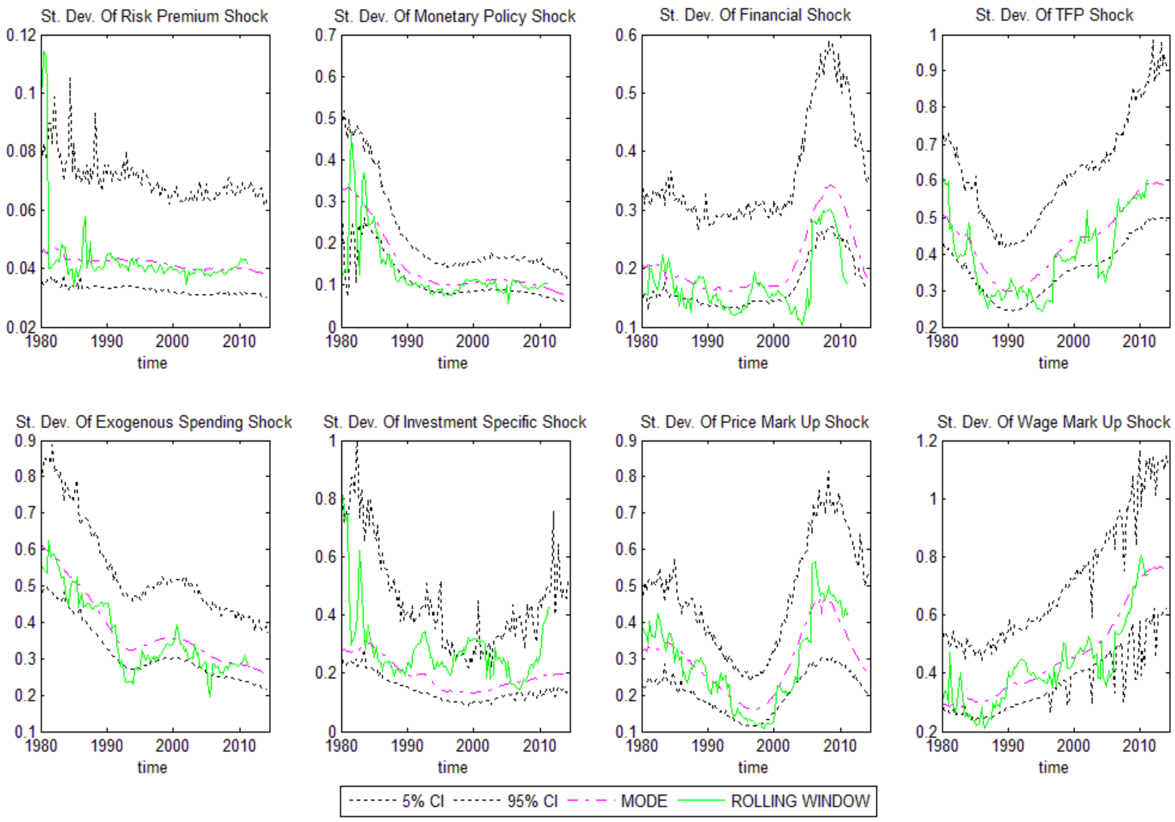


Figure 54: Additional Robustness Checks

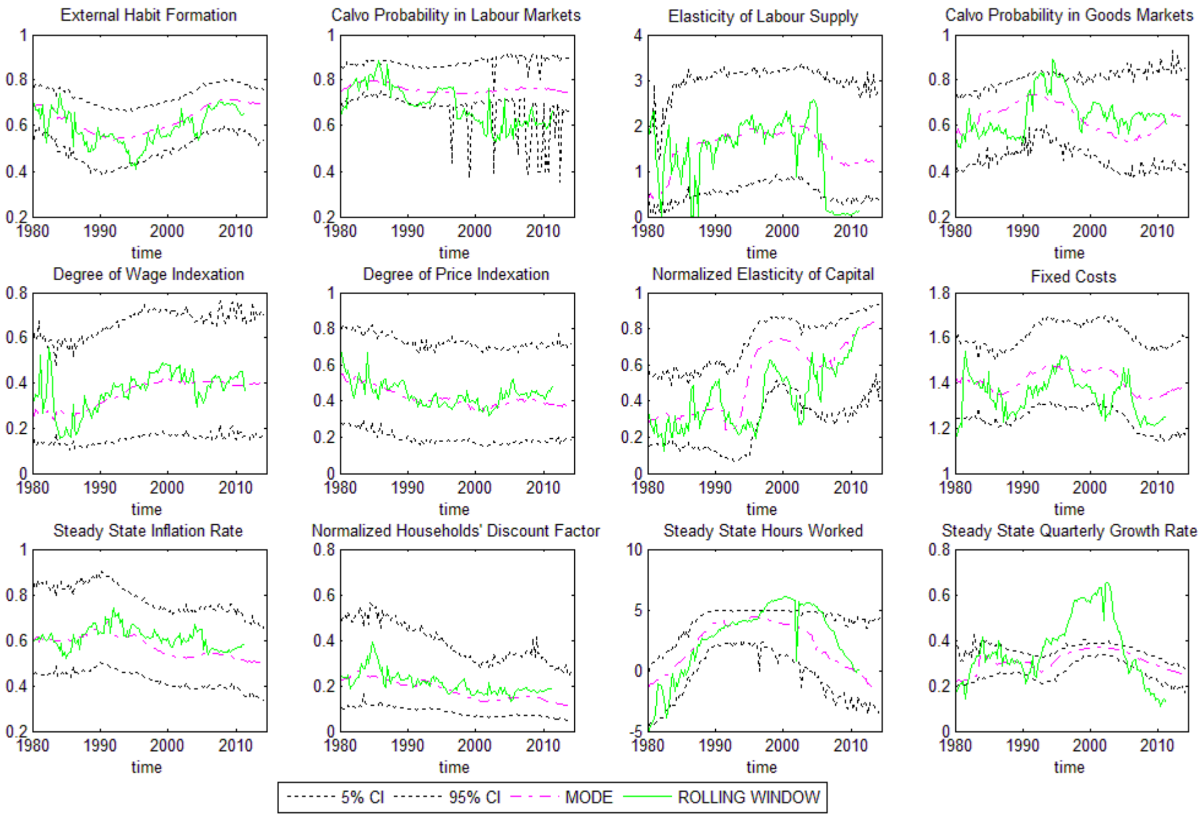
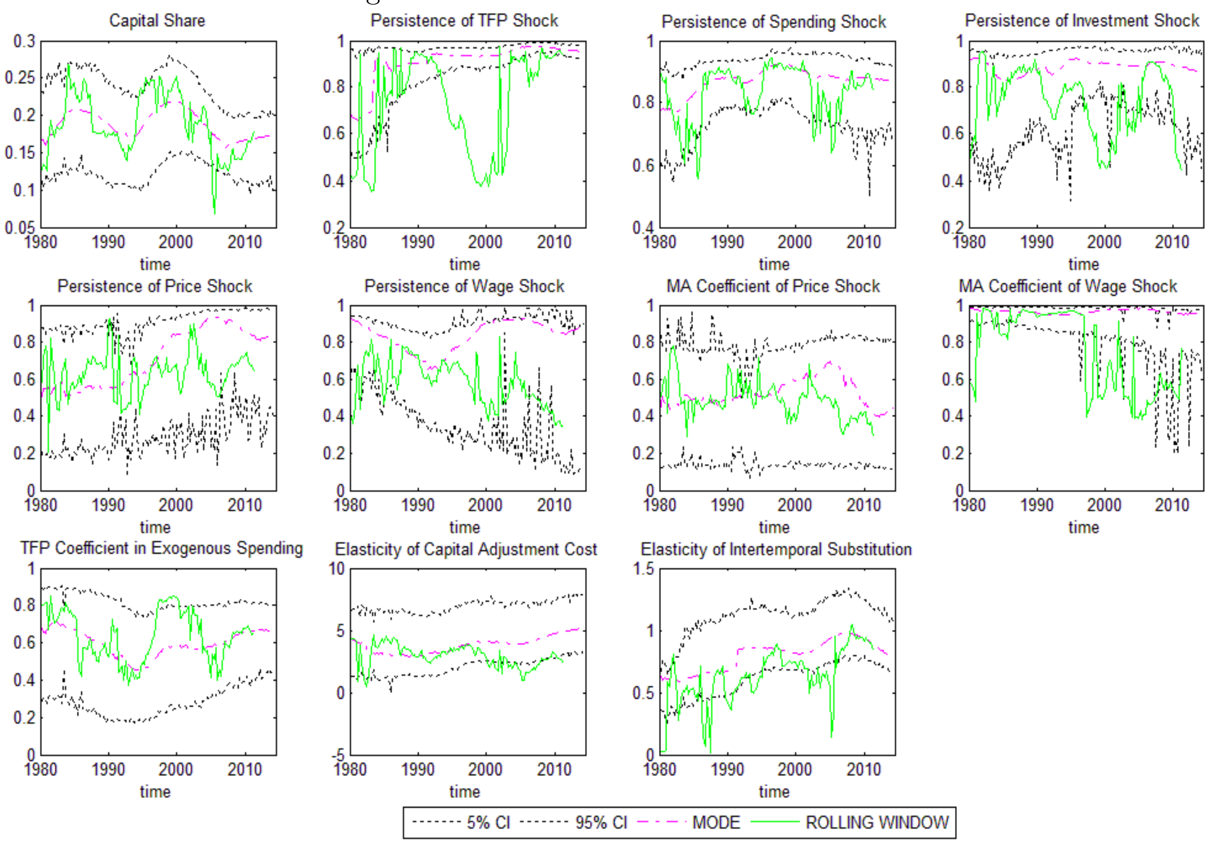


Figure 55: Additional Robustness Checks



7.3.4 Robustness check: different spread variable

Figures 56, 57 and 58 display the additional robustness checks with respect to the spread variable for Chapter 4. The posterior mode obtained by QBLL with spread BAA corporate bond yield over 10 year Treasury note is the pink dash-dotted line, the 5% and 95% posterior quantile values are the black dotted lines, the posterior mode obtained by QBLL with spread BAA corporate bond yield minus Fed Funds Rate is the solid blue line.

Figure 56: Additional Robustness Checks

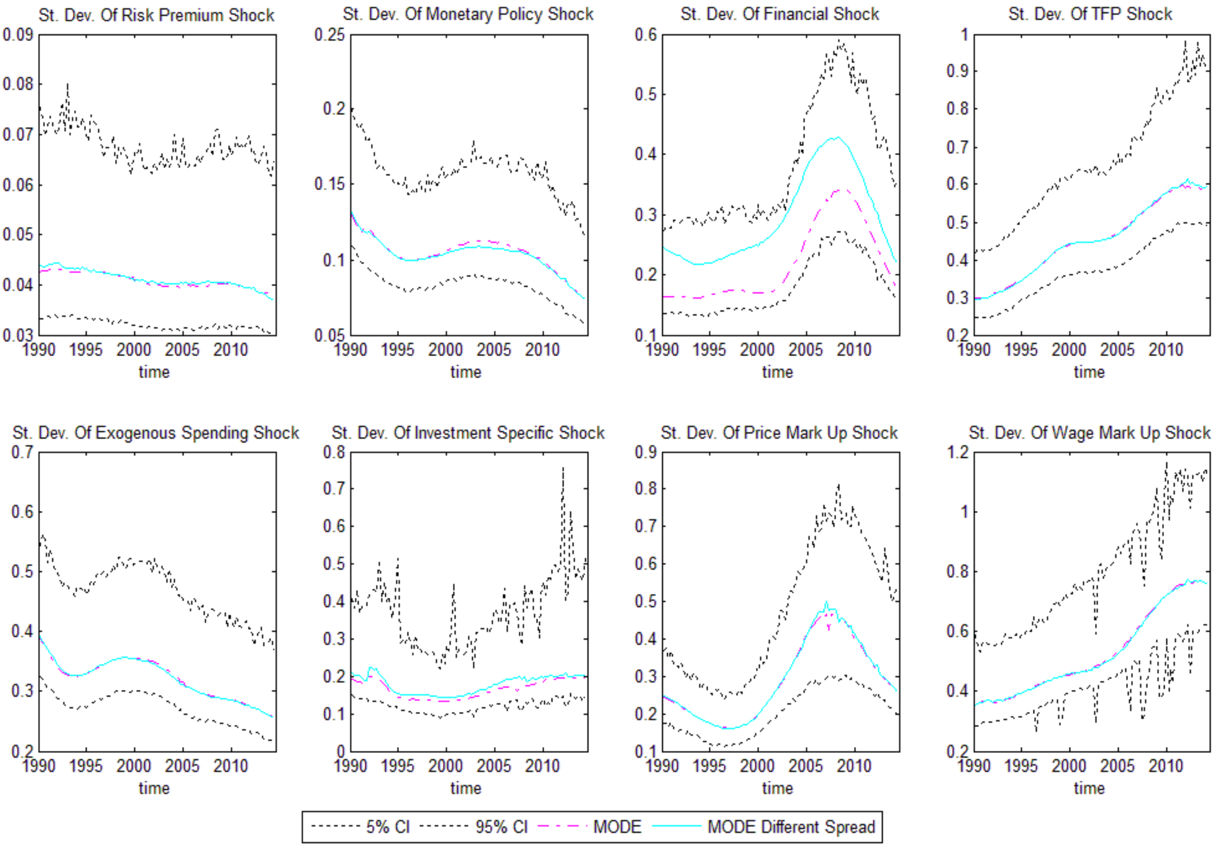


Figure 57: Additional Robustness Checks

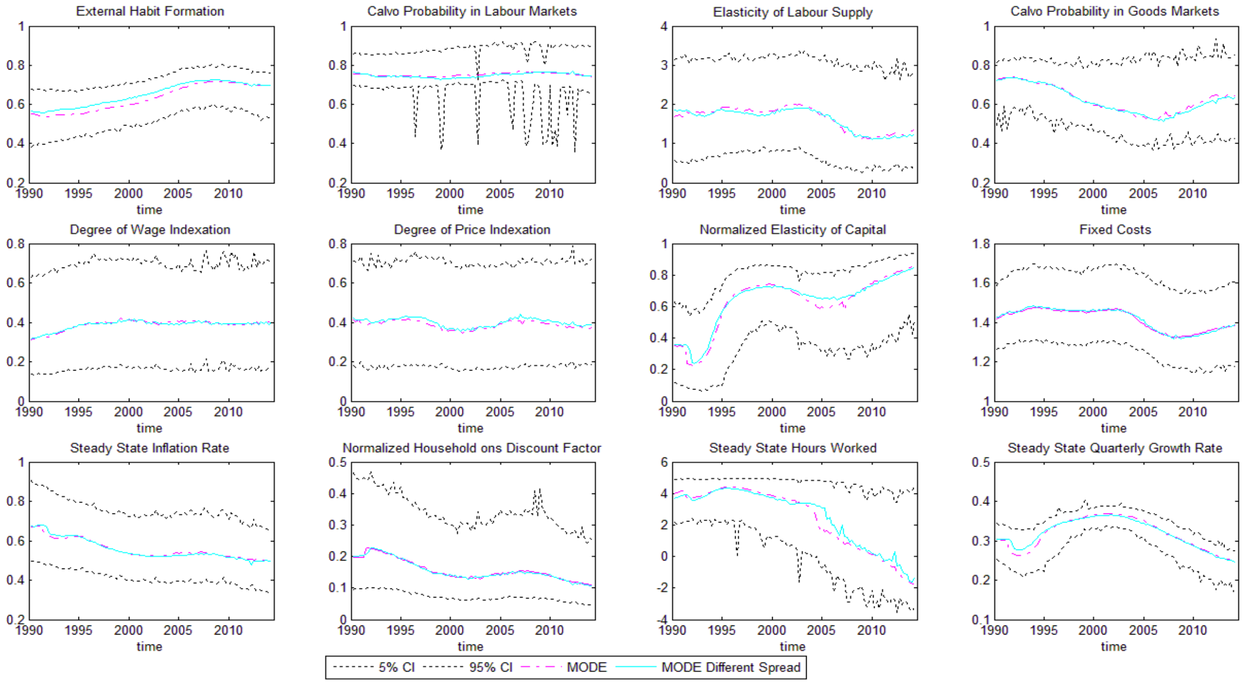
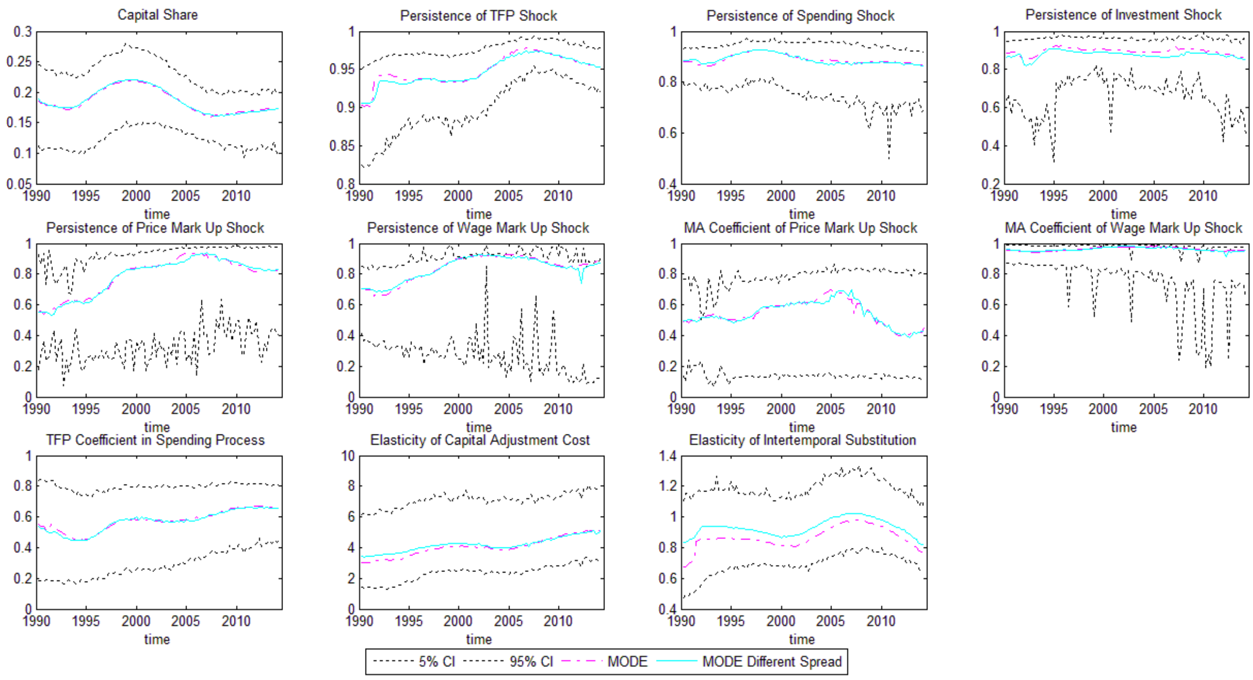


Figure 58: Additional Robustness Checks



7.3.5 Robustness check: Simulation Exercise

Figures 59, 60 and 61 contain the additional simulation exercise results for Chapter 4. The posterior mode obtained by QBLL when the DGP is a model with fixed parameters is the solid blue line and the true parameter values at which the data are generated is the dotted green line.

Figure 59: Additional Simulation Exercise Results

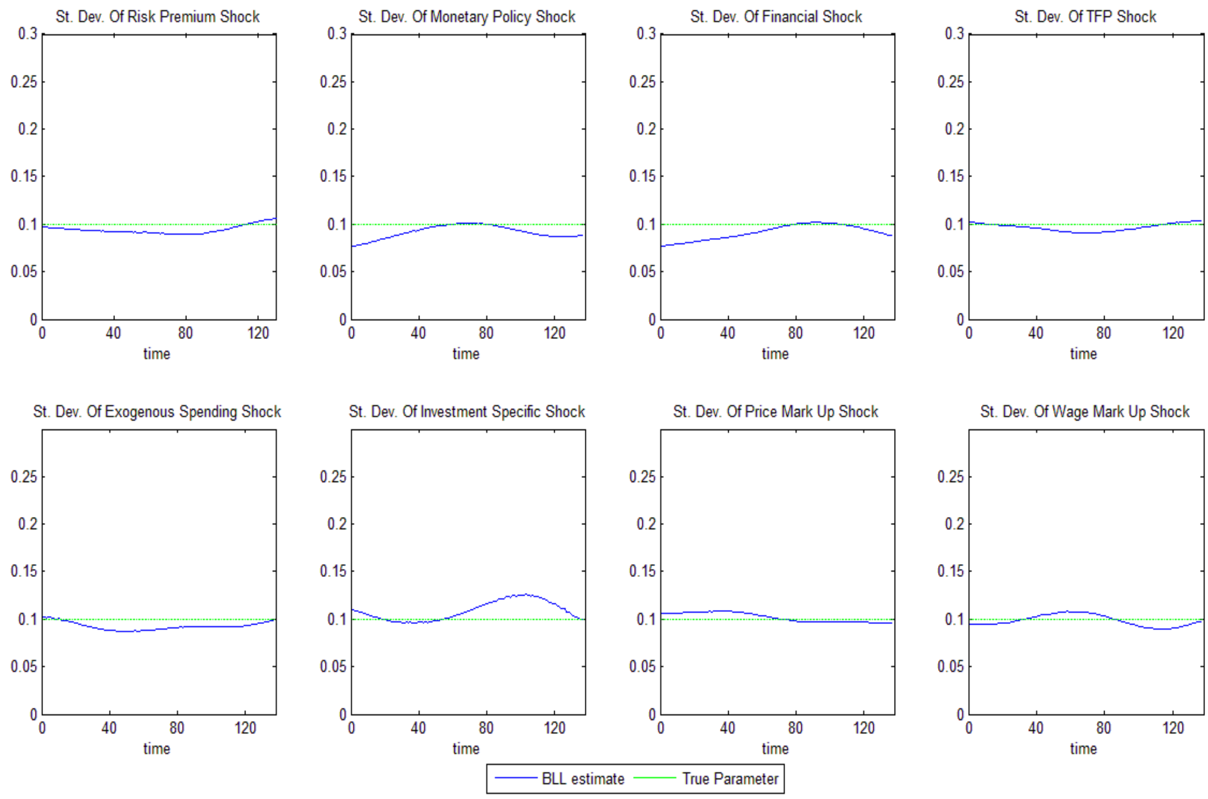


Figure 60: Additional Simulation Exercise Results

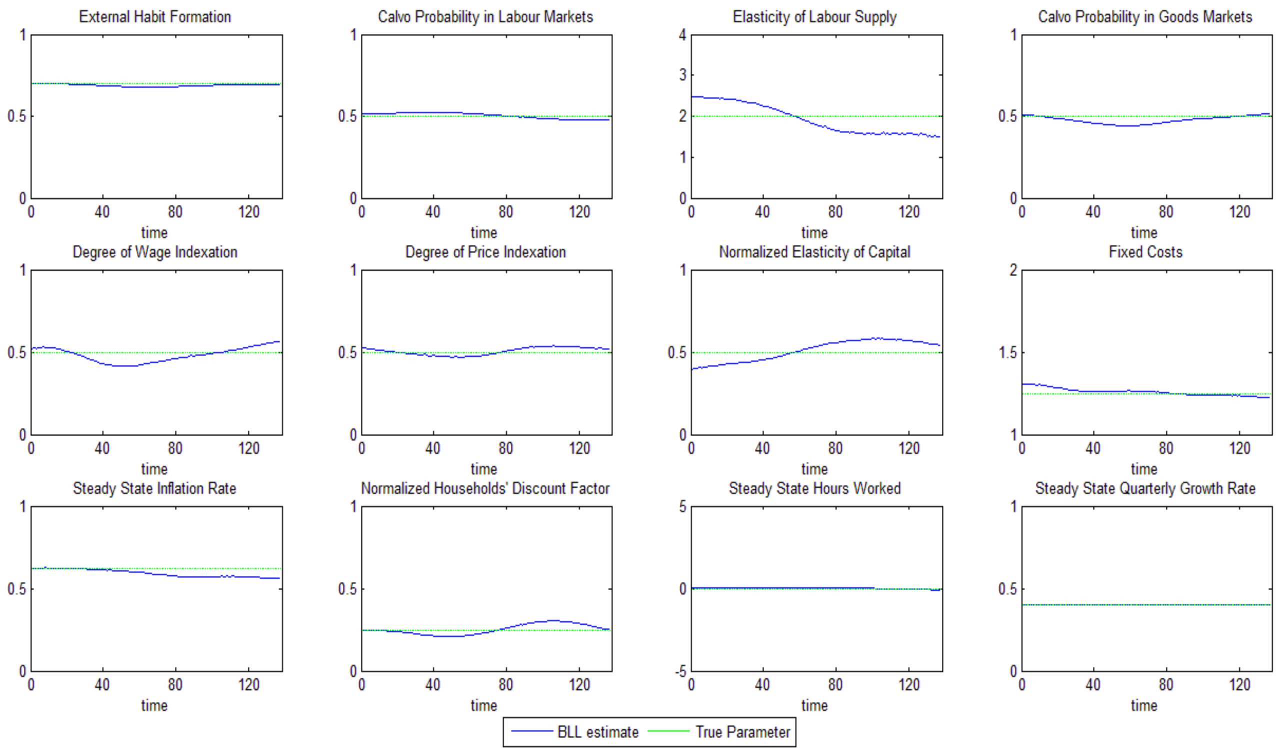
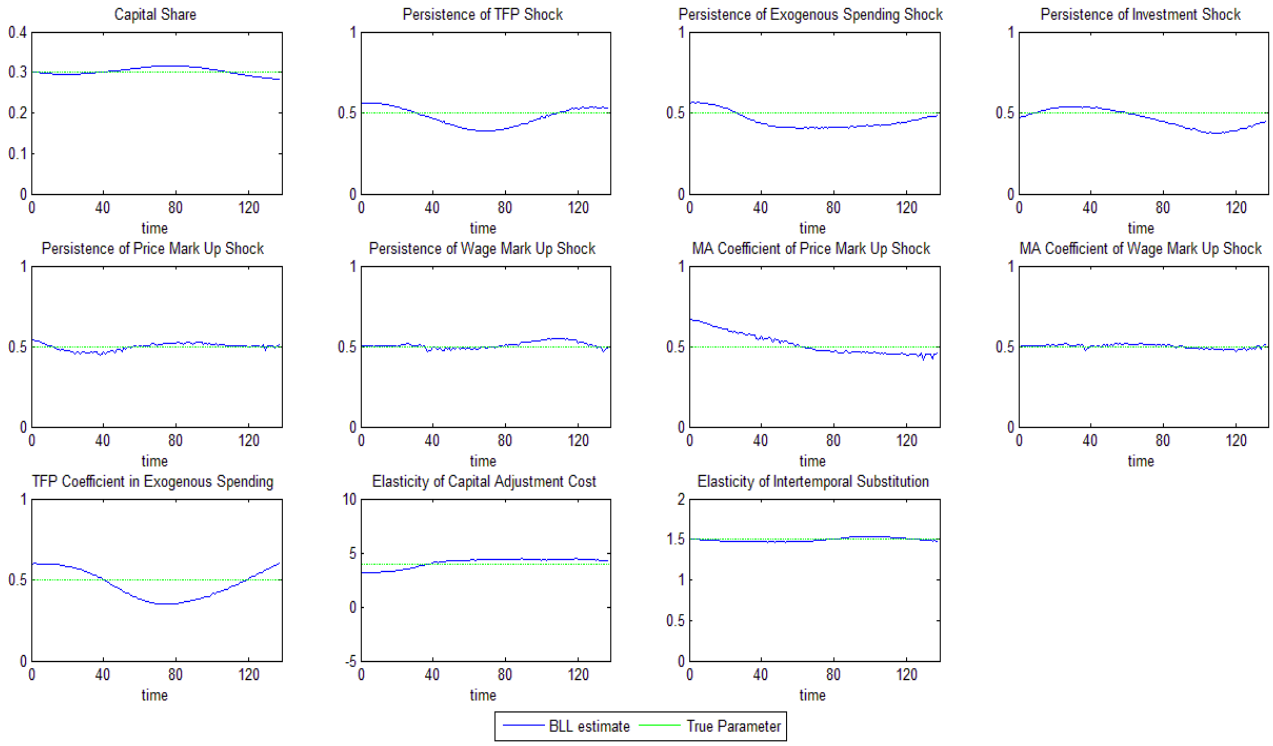


Figure 61: Additional Simulation Exercise Results



7.4 Model Data Descriptions and Additional Results for Chapter 5

Parameter	Description	Prior Distribution	Prior Mean	Prior Std
θ_R	Policy rule interest rate smoothing	Beta	0.800	0.100
θ_Π	Policy rule inflation response	Normal	1.500	0.250
θ_Y	Policy rule output gap response	Beta	0.125	0.075
ϕ_Z	Final output price adjustment cost	Gamma	7.000	2.000
ϕ_V	Value added price adjustment cost	Gamma	7.000	2.000
ϕ_M	Import price adjustment cost	Gamma	10.00	2.000
ϕ_X	Export price adjustment cost	Gamma	10.00	2.000
ϕ_W	Nominal wage adjustment cost	Gamma	14.00	2.000
ξ_Z	Indexation of final output prices	Beta	0.25	0.075
ξ_V	Indexation of value added prices	Beta	0.25	0.075
ξ_M	Indexation of import prices	Beta	0.25	0.075
ξ_X	Indexation of export prices	Beta	0.25	0.075
ξ_W	Indexation of nominal wages	Beta	0.25	0.075
ψ_C	Habit formation parameter	Beta	0.70	0.150
ψ_I	Investment adjustment cost	Gamma	2.00	0.400
ϵ_C	Coefficient of relative risk aversion	Gamma	1.50	0.200
ϵ_L	Labour supply elasticity	Gamma	2.00	0.300
ϵ_F	Price elasticity world demand, UK exports	Gamma	0.75	0.100
ω_o	Share of optimising households	Beta	0.70	0.050
ρ_B	Persistence of risk premium forcing process	Beta	0.75	0.100
ρ_I	Persistence of investment adjustment shock	Beta	0.75	0.100
ρ_G	Persistence of government spending shock	Beta	0.90	0.050
ρ_{IO}	Persistence of other investment shock	Beta	0.75	0.100
$\rho_{\kappa F}$	Persistence of export preference shock	Beta	0.75	0.100
ρ_M	Persistence of import preference shock	Beta	0.75	0.100
ρ_L	Persistence of labour supply shock	Beta	0.75	0.100

Table 33. Priors for estimated parameters.

Parameter	Description	Prior Distribution	Prior Mean	Prior Std
ρ_{B^F}	Persistence of UIP shock	Beta	0.75	0.10
ρ_{PX^F}	Persistence of world export price shock	Beta	0.90	0.05
ρ_{Z^F}	Persistence of world output shock	Beta	0.90	0.05
σ_B	St dev of risk premium shock	Gamma	0.50	0.20
σ_I	St dev of investment adjustment shock	Gamma	1.90	0.20
σ_G	St dev of government spending shock	Gamma	3.00	0.20
σ_{IO}	St dev of other investment shock	Gamma	14.0	1.00
σ_{κ^F}	St dev of export preference shock	Gamma	2.20	0.20
σ_M	St dev of import preference shock	Gamma	2.20	0.20
σ_{LAP}	St dev of LAP growth shock	Gamma	0.35	0.10
σ_L	St dev of labour supply shock	Gamma	0.75	0.20
σ_R	St dev of monetary policy shock	Gamma	0.10	0.10
σ_{B^F}	St dev of UIP shock	Gamma	0.65	0.20
σ_{μ^Z}	St dev of final output markup shock	Gamma	0.10	0.10
σ_{μ^W}	St dev of wage markup shock	Gamma	0.30	0.10
σ_{μ^M}	St dev of import markup shock	Gamma	1.30	0.20
σ_{μ^X}	St dev of export markup shock	Gamma	1.30	0.20
σ_{PX^F}	St dev of world export price shock	Gamma	1.60	0.20
σ_{Z^F}	St dev of world output shock	Gamma	2.50	0.20
σ_i^{me}	St dev of investment measurement error	Gamma	0.35	0.10
σ_X^{me}	St dev of export measurement error	Gamma	0.18	0.055
σ_M^{me}	St dev of import measurement error	Gamma	0.18	0.055
σ_L^{me}	St dev of hours measurement error	Gamma	0.045	0.013
σ_W^{me}	St dev of wage measurement error	Gamma	0.125	0.0275
σ_{PM}^{me}	St dev of import price measurement error	Gamma	0.34	0.075
σ_{PX}^{me}	St dev of export price measurement error	Gamma	0.34	0.075

Table 34. Priors for estimated parameters.

Variable	Description	Data transformation equation	Measurement equation
gdpkp	Real GDP	$\mathrm{dlngdpkp}_t \equiv 100\Delta \ln \mathrm{gdpkp}_t$	$\Delta v_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H \left(\Gamma^X\right)^{-\frac{1-\alpha_V}{\alpha_V}} \right)$
ckp	Real cons.	$\mathrm{dlnckp}_t \equiv 100\Delta \ln \mathrm{ckp}_t$	$\Delta c_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H\right)$
ikkp	Real inv.	$\mathrm{dlnikkp}_t \equiv 100\Delta \ln \mathrm{ikkp}_t$	$\Delta i_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H \Gamma^I\right) + \sigma_I^{me} me_t^I$
gonskp	Real spending	$\mathrm{dlngonskp}_t \equiv 100\Delta \ln \mathrm{gonskp}_t$	$\Delta g_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H \Gamma^G\right)$
xkp	Real exports	$\mathrm{dlnxkp}_t \equiv 100\Delta \ln \mathrm{xkp}_t - \mathrm{dlnxkp}_t^{\mathrm{tt}}$	$\Delta x_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H \Gamma^X\right) + \sigma_X^{me} me_t^X$
mkp	Real imports	$\mathrm{dlnmkp}_t \equiv 100\Delta \ln \mathrm{mkp}_t - \mathrm{dlnmkp}_t^{\mathrm{tt}}$	$\Delta m_t + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H \Gamma^X\right) + \sigma_M^{me} me_t^M$
pxdef	Export deflator	$\mathrm{dlnpxdef}_t \equiv 100\Delta \ln \mathrm{pxdef}_t - \Pi_t^{*,\mathrm{tt}} - \Pi_t^{x,\mathrm{tt}}$	$\Delta p_t^{EX} - \Delta q_t + \pi_t^Z + 100 \ln \frac{\Pi^*}{\Gamma^X} + \sigma_{PX}^{me} me_t^{PX}$
pmdef	Import deflator	$\mathrm{dlnpmdef}_t \equiv 100\Delta \ln \mathrm{pmdef}_t - \Pi_t^{*,\mathrm{tt}} - \Pi_t^{m,\mathrm{tt}}$	$\pi_t^M + 100 \ln \frac{\Pi^*}{\Gamma^X} + \sigma_{PM}^{me} me_t^{PM}$
awe	Nom. wage	$\mathrm{dlnawe}_t \equiv \Delta \ln \mathrm{awe}_t - \Pi_t^{*,\mathrm{tt}}$	$\Delta w_t + \gamma_t^Z + \pi_t^Z + 100 \ln \left(\Gamma^Z \Pi^*\right) + \sigma_W^{me} me_t^W$
cpisa	SA CPI	$\mathrm{dlncpisa}_t \equiv 100\Delta \ln \mathrm{cpisa}_t - \Pi_t^{*,\mathrm{tt}}$	$\pi_t^C + 100 \ln \Pi^*$
rga	Bank Rate	$\mathrm{robs}_t \equiv 100 \ln \left(1 + \frac{\mathrm{rga}_t}{100}\right)^{\frac{1}{4}} - \Pi_t^{*,\mathrm{tt}}$	$r_t + 100 \ln R$
eer	Sterling ERI	$\mathrm{dlneer}_t \equiv 100\Delta \ln \mathrm{eer}_t$	$\Delta q_t - \pi_t^Z$
hrs	Hours worked	$\mathrm{dlnhrs}_t \equiv 100\Delta \ln \mathrm{hrs}_t$	$\Delta l_t + 100 \ln \Gamma^H + \sigma_L^{me} me_t^L$
yf	World output	$\mathrm{dlnyf}_t \equiv 100\Delta \ln \mathrm{yf}_t - \mathrm{dlnyf}_t^{\mathrm{tt}}$	$\Delta z_t^F + \gamma_t^Z + 100 \ln \left(\Gamma^Z \Gamma^H\right)$
pxfdef	World exp. def.	$\mathrm{dlnpxfdef}_t \equiv 100\Delta \ln \mathrm{pxfdef}_t - \Pi_t^{\mathrm{xf},\mathrm{tt}}$	$\Delta p_t^{X^F} + 100 \ln \frac{\Pi^*}{\Gamma^X}$

Table 35. Observables, data transformation and measurement equations.

7.5 Examples of parameter processes

In a referee report, we were asked to be more explicit about what parameter processes the approach can recover. We include a brief discussion below. Recall that we require the parameters to satisfy one of the following:

(i) For each $t \in \{1, ..., T\}$, θ_t is a deterministic function of time given by

$$\theta_t = \theta\left(\frac{t}{T}\right)$$

where $\theta(.)$ is a piecewise differentiable function.

(ii) For $1 \leq h \leq t$ as $h \to \infty$, θ_t is a vector-valued stochastic process satisfying

$$\sup_{j:|j-t|\leq h} ||\theta_t - \theta_j||^2 = O_p\left(h/t\right).$$

Condition i) covers deterministic functions of the time fraction t/T, constant parameters as a special case as well as breaks in the parameter processes. Condition ii) allows the process to be stochastic but requires it to be persistent in order for the approach to be able to recover the time variation. Note that ii) includes processes such as the bounded random walk process used in DGPII and DGPIII in Chapter 2:

$$\theta_t = \frac{\sum_{i=1}^t \varepsilon_t}{\sqrt{t}}, \quad \varepsilon_t \sim \mathcal{N}(0, v).$$

This process can equivalently be written as

$$\theta_t = \sqrt{\frac{t-1}{t}}\theta_{t-1} + \frac{\varepsilon_t}{\sqrt{t}},$$

so that the variance of the innovations is decreasing while the process is becoming closer to unit root, and hence the variance of the process θ_t stabilises. In addition, condition ii) also covers long memory processes with memory parameter $d > 0.5$. A stationary AR(1) process is not covered in condition ii), as it is not persistent enough. Finally, the approach developed in this thesis is also valid for a combination processes between i) and ii). In the figure below, we provide several examples of parameter processes satisfying condition i) or ii) or a combination of the two.

Figure 62: Examples of parameter processes satisfying i) or ii)